Thesis for the Degree of Ph.D.

# SOME INVESTIGATIONS IN STATISTICAL FREQUENCY OF TYPE B

by

J.T.CAMPBELL, M.A. (N.Z.)

---------------

Degree conferred 30 June, 1932



EDINBURGH.

April, 1932.

## INTRODUCTION.

The most common types of statistical distribution involve variables $x$ which are regarded as the resultant, or integral, of elementary variables or increments either discrete, as $\Delta x$, or in the shape of infinitesimal differentials, $dx$. The elementary variables have their own separate laws of distribution, and the distribution of the resultant variable $x$ depends on the nature of these laws. For example, if the elementary increments are uncorrelated, and if the probability of non-zero increments is not of too small an order, then the normal frequency function

$$dp = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}x^2/\sigma^2} dx = \phi(x)\,dx,$$

can be derived under very wide hypotheses, as the frequency function of the resultant variable $x$, while if the increments are discrete and finite in number, a closer approximation is given by the function called by Charlier "Type A", consisting of the normal function $\phi(x)$ supplemented by corrective terms involving its third and higher derivatives.

If the probability of non-zero increments $\Delta x$ is of <u>smaller</u> order than in the above case, more precisely of order $1/N$ where $N$ is the number of increments, the appropriate frequency function of $x$ proves to be the Skew function of Poisson,

$$p = e^{-a}a^x/x! = \psi(x),$$

a closer approximation being the Charlier series of "Type B" which consists of the Poisson function $\psi(x)$ supplemented by terms involving the second and higher "receding" finite differences (i.e. defined by $f(x) - f(x-1)$ ) of $\psi(x)$.

Sufficient mathematical conditions for the derivation of Types A and B have occupied the attention of recent investigators (9; 4)[1); but it is not easy in practical work to verify that these conditions have been present.

In what follows we shall attempt to develop the theory of the Poisson function and the function of Type B, in the direction of extending them to the case of two or more correlated variables. We shall find that the Poisson function is not so amenable to such extension as the normal frequency function. For example, the normal frequency distribution of several linearly correlated variables simply involves the exponential of a negative quadratic form, which in the non-correlated case is a sum of squares, and free of product terms; but in the Poisson case we could find no such simple explicit form for the multiple Poisson function of correlated variables. There is however, an alternative form for the normal correlation function for two variables,

$$\phi(x, y) = \frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{1}{2}(x^2 - 2rxy + y^2)/(1-r^2)}:$$

it is (11.p.224)

$$\phi(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \left(1 + H_1(x) H_1(y) r + H_2(x) H_2(y) \frac{r^2}{2!} + \cdots\right),$$

where $H_r(x)$ stands for the familiar orthogonal polynomials of Hermite. We shall adopt this as the representative form of the

1) See bibliography.

correlation function, and shall develop an analogous form in the Poisson case. The parallelism between this and the normal case will be emphasized by the existence of orthogonal polynomials with properties resembling those of the Hermite polynomials.

The introduction in § 5 of factorial moments and the factorial moment generating function will lead in subsequent paragraphs to results and properties for frequencies of Type B, that show remarkable analogies to those discussed earlier in Chapter I for frequencies of Type A.

Of particular interest in Chapter II is the relation (§8) found to exist for the distances between the mean, median and mode of the Poisson distribution.

In the last section an account will be given of an experiment undertaken to verify the theoretical formulae of double Poisson correlation. We were unable to find any statistical records of correlation of two variables each obeying a Poisson law of distribution, and therefore had to devise the artificial experiment with marbles described.

§§ 5, 6, 10, 11, 12 and 15 are from a paper that has been submitted for publication in the Proceedings of the Edinburgh Mathematical Society.

Finally I desire to express my very deep indebtedness to Dr.A.C.Aitken of the Mathematical Department, Edinburgh, for his continued encouragement and guidance.

# CHAPTER I.

## MOMENT GENERATING FUNCTIONS AND FACTORIAL MOMENT GENERATING FUNCTIONS.

### §1. GENERATING FUNCTIONS.

In the study of frequency distributions and the distributions of the derived parameters, a very powerful method of attack introduced by Laplace is to be found in the employment of generating functions. If $\phi(x)$ is a function of $x$, then

$$F(t) = \sum_{x=-\infty}^{\infty} \phi(x) t^x , \qquad \text{or} \qquad F(t) = \int_{-\infty}^{\infty} \phi(x) t^x dx,$$

according as $x$ varies discontinuously or continuously, is defined as the generating function of $\phi(x)$. In most of the cases that will be considered here $\phi(x)$ will represent a frequency, so that $\sum_{x=-\infty}^{\infty} \phi(x)$ or $\int_{-\infty}^{\infty} \phi(x) dx$ is unity.

If the generating function is multiplied by $t^{-m}$ or $t^m$ where $m$ is a constant, the effect is to reduce or to increase all the indices of $t$ in the generating function, by $m$, i.e. to advance or put back the origin of measurement.

Further, if $t^s$ is replaced by $t_1^{s_1}$, the result is to change the unit of measurement, such that $s_1$ units in the new scale are equivalent to $s$ units in the old.

As examples of such generating functions there are the following:-

(i) $(pt+q)^n$ where $p+q=1$. This is the generating function of the relative frequencies of $n$, $n-1$, $n-2$, ...... occurrences of a character A in samples of $n$, drawn from an infinite population consisting of A and not -A in the proportion $p:q$. It is also the generating function of samples of $n$ from a finite but similar population when there are replacements after each selection.

This can be extended to the case of sampling from a population consisting of several characters which are in the proportion $p_1 : p_2 : p_3 : \dots$, where $p_1 + p_2 + p_3 + \dots = 1$, when the generating function of the frequencies in samples of $n$ is given by

$$(p_1 t_1 + p_2 t_2 + p_3 t_3 + \dots)^n.$$

By the introduction of the variables t we are able to keep account of the number of appearances of any particular character.

(ii) An example of a generating function which is of interest in pure mathematics as well as in statistics, is

$$e^{-\frac{1}{2}t^2 + tx},$$

which is the generating function of the Hermite polynomial, or parabolic cylinder function, that of order $s$, being defined by

$$H_s(x) = (-)^s e^{\frac{1}{2}x^2} \frac{d^s}{dx^s}\left(e^{-\frac{1}{2}x^2}\right)$$

$$= x^s - \frac{s(s-1)}{2 \cdot 1!} x^{s-2} + \frac{s(s-1)(s-2)(s-3)}{2^2 \cdot 2!} x^{s-4} \dots .$$

## §2. MOMENT GENERATING FUNCTIONS.

If now in the generating function

$$\Sigma f(x) t^x,$$

$t$ is replaced by $e^{\alpha}$ then

$$F(e^{\alpha}) = \sum \phi(x) e^{\alpha x} \quad \text{------------} (2.1)$$

$$= \sum \phi(x) + \alpha \sum x \phi(x) + \frac{\alpha^2}{2!} \sum x^2 \phi(x) + \cdots,$$

where the coefficient of $\alpha^r/r!$ is $\sum x^r \phi(x)$, i.e. the $r^{th}$ moment of the function $\phi(x)$. The expression on the right hand side of (2.1) is called the moment generating function of $\phi(x)$. Hence it is seen that the moment generating function of $\phi(x)$ is obtained simply from the generating function of $\phi(x)$ by replacing $t$ in the latter by $e^{\alpha}$. It should be noted that the $r^{th}$ moment is the coefficient of $\alpha^r/r!$, not just of $\alpha^r$, in the moment generating function.

From what was stated in §1 it is clear that to change the origin of measurement forwards or backwards by an amount $m$ is to multiply the moment generating function by $e^{-m\alpha}$ or $e^{m\alpha}$, and to change the scale of measurement in the ratio $s:s_1$, is to replace $s\alpha$ by $s_1\alpha_1$.

Thus the moment generating function of the binomial distribution given above is

$$(pe^{\alpha}+q)^n,$$

$$\text{or} \quad (1 + p\alpha + p\alpha^2/2! + \cdots)^n,$$

whence we have at once that the mean is $np$, and further moments of the distribution can be obtained by picking out the coefficient of the required power of $\alpha$. If now, the moments of the distribution are desired referred to the mean as origin, the required moment generating function is

$$e^{-np\alpha}(pe^{\alpha}+q)^n,$$

or $$(pe^{q\alpha}+qe^{-p\alpha})^n$$

$$= (1+ pq\alpha^2/2! + pq(q-p)\alpha^3/3! + \cdots)^n$$

therefore the variance, or squared standard deviation is $npq$, and the third moment $npq(q-p)$.

The following are further examples of moment generating functions:-

(i) $e^{\frac{1}{2}\alpha^2}$ : which is the moment generating function of

$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$, the Gaussian normal function;

(ii) $e^{\frac{1}{2}\alpha^2+\frac{1}{2}\beta^2+r\alpha\beta}$ : the moment generating function of

$$\phi(x,y) = \frac{1}{2\pi\sqrt{1-r^2}}e^{-\frac{1}{2}(x^2-2rxy+y^2)/(1-r^2)}$$, the normal

correlation function in two variables. These expressions are obtained as the limiting cases, as $N \to \infty$, of the moment generating function of samples of $N$ drawn from Bernoullian or Poisson populations involving for (i) one character, and for (ii) two correlated characters. They also arise under more general conditions when the independent variables are considered to be compounded from a large number of elementary increments.

The moment generating function having thus been found, the frequency function is obtained by solving the integral equation arising from equating the two forms of the moment generating function, for example in the case of (i),

$$e^{\frac{1}{2}\alpha^2} = \int_{-\infty}^{\infty}\phi(x)e^{\alpha x}dx,$$

an integral equation for the function $\phi(x)$, the solution of which gives the normal function. The normal correlation function in $n$ variables can be similarly obtained as the solution of an integral equation (7).

## §3. EXPANSION OF THE CHARLIER TYPE A FUNCTION IN A SERIES OF HERMITE POLYNOMIALS.

It can be shown (1) that if $J(f)$ is the moment generating function of an arbitrary function $f(x)$ and we put

$$J(f)e^{-\frac{1}{2}\alpha^2} = \sum_{n=0}^{\infty} A_n \alpha^n,$$

then,

$$A_n = \frac{(-)^n}{n!}\int_{-\infty}^{\infty} f(x)\left(x^n - \frac{n(n-1)}{2.1!}x^{n-2} + \cdots\right)dx$$

$$= \frac{(-)^n}{n!}\int_{-\infty}^{\infty} f(x)\, H_n(x)\, dx;$$

and further, if $f(x)$ is such that one of the derivatives of $f(x)e^{\frac{1}{2}x^2}$ has total limited fluctuation, then $f(x)$ can be expressed as an absolutely and uniformly convergent series of Hermite polynomials of the form,

$$f(x) = \sum_{n=0}^{\infty} A_n \frac{d^n}{dx^n}\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)$$

$$= \sum_{n=0}^{\infty} (-)^n A_n . H_n(x) . \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

where the coefficient $A_n$ is that given above.

The moment generating function of the Charlier Type A function, $f(x)$ say, the generalized form of the normal function, is

$$e^{\frac{1}{2}\alpha^2}\left(1+a_3\alpha^3+a_4\alpha^4+\cdots\right),$$

in which the $a$'s are parameters and are of descending order in N, where N is the number of elementary increments considered as contributing to any value $x$ of the independent variable of the distribution. The normal function is obtained as the limiting case of this when terms of order $1/\sqrt{N}$ and higher are neglected. When we retain these higher terms and apply the theorem given above we obtain the familiar Charlier Type A expansion (8),

$$f(x) = \phi(x) + a_3\frac{d^3}{dx^3}\phi(x) + a_4\frac{d^4}{dx^4}\phi(x) + \cdots,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ and $a_n = A_n = \frac{(-)^n}{n!}\int_{-\infty}^{\infty} f(x)H_n(x)\,dx.$

An interesting discussion of this expansion regarded as an asymptotic series will be found in H.Cramer's Memoir (9).

The coefficient $a_n$ can be obtained from the last expression in terms of the moments of the frequency function $f(x)$, by writing out $H_n(x)$ explicitly in terms of $x$, whence

$$a_n = \frac{(-)^n}{n!}\left(\mu_n - \frac{n(n-1)}{2.1!}\mu_{n-2} + \frac{n(n-1)(n-2)(n-3)}{2^2.2!}\mu_{n-4} - \cdots\right),$$

where $\mu_r = \int_{-\infty}^{\infty} x^r f(x)\,dx$. The particular form of the Type A function, with the coefficients $a_1$ and $a_2$ missing, arises through the mean of the distribution being taken as origin and the independent variable being expressed in terms of the standard deviation as unit, i.e., the variables are 'normalized'. The values of the first two coefficients are thus:

$$a_3 = -\frac{1}{3!}(\mu_3 - 3\mu_1)$$
$$= -\mu_3/3!, \text{ as } \mu_1 = 0,$$

and

$$a_4 = \frac{1}{4!}(\mu_4 - 6\mu_2 + 3)$$
$$= \frac{1}{4!}(\mu_4 - 3), \text{ as } \mu_2 = 1.$$

These relations between the $a$'s and $\mu$'s are required when a curve of Type A is being fitted to given data.

The mistake is frequently made in applying the Type A formula, of fitting the given data as far as the term in $a_4$. Strictly, if one wishes to take in the term in $a_4$ that in $a_6$ should be included as well. This is due to the fact that the $a$'s do not decrease regularly in order with respect to $N$, but in groups of the form $a_3$, $a_4$ and $a_6$, $a_5$, $a_7$ and $a_9$, and so on.(10).

## §4. THE NORMAL CORRELATION FUNCTION AS A POWER SERIES IN $r$.

Consideration of the moment generating function of the normal correlation function in two variables leads to an elegant method of expressing this frequency function as a power series in $r$, the correlation coefficient.(11.p.224). This result depends on a theorem (1.VII.) on generating functions that, if $F(\alpha)$ is the moment generating function of $\phi(x)$ then $\alpha^s F(\alpha)$ is the moment generating function of $\left(-\frac{d}{dx}\right)^s \phi(x)$. This result is also true for moment generating functions involving several variables, the total derivatives then being replaced by partial derivatives.

The moment generating function of two normally correlated variables is $e^{\frac{1}{2}\alpha^2+\frac{1}{2}\beta^2+r\alpha\beta}$ and that of two normal and uncorrelated variables $e^{\frac{1}{2}\alpha^2+\frac{1}{2}\beta^2}$. Hence applying the result stated above, since by that $\alpha$ is equivalent to the operation $-\frac{\partial}{\partial x}$ and $\beta$ to the operation $-\frac{\partial}{\partial y}$, we obtain the operational result,

$$\phi(x,y) = \frac{1}{2\pi}\, e^{r\left(-\frac{\partial}{\partial x}\right)\left(-\frac{\partial}{\partial y}\right)} \cdot e^{-\frac{1}{2}(x^2+y^2)}$$

$$= \frac{1}{2\pi}\, e^{-\frac{1}{2}(x^2+y^2)} \left(1+H_1(x)H_1(y)\cdot r+H_2(x)H_2(y)\cdot\frac{r^2}{2!}+\cdots\right),$$

where $\phi(x,y)$ is the normal correlation function in two variables. This expansion is used in obtaining the moments of any section of the normal correlation surface.

The above result can be extended at once to the normal correlation function in $n$ variables (6.p.21), when we shall have

$$\phi(x_1,x_2,\ldots x_n) = \frac{1}{(2\pi)^{n/2}}\, e^{\sum_{s,t=1}^{n} r_{s,t}\left(-\frac{\partial}{\partial x_s}\right)\left(-\frac{\partial}{\partial x_t}\right)} \cdot e^{-\frac{1}{2}\sum_{s=1}^{n} x_s^2}$$

where $\phi(x_1,x_2,\ldots x_n) = \frac{1}{(2\pi)^{n/2}|R|^{1/2}}\, e^{-\frac{1}{2}R^{-1}(x,x)}$,

$R^{-1}(x,x)$ being the reciprocal quadratic form of the quadratic form $R(x,x)=\sum_{s,t=1}^{n} r_{st}x_s x_t$, whose determinant is denoted by $|R|$.

## §5. FACTORIAL MOMENTS AND FACTORIAL MOMENT GENERATING FUNCTIONS.

When a given frequency distribution is to be graduated it is customary, as we did in §3, to express the parameters of the fitted curve in terms of the moments of the frequency distribution. The $r^{th}$ moment of the distribution in which the relative frequency of a measure $x$ is $\phi(x)$, or in the case of a continuous variable, is given by the differential of frequency $\phi(x)dx$, is defined in the respective cases by

$$\sum_x x^r \phi(x) \quad , \text{ or by } \int x^r \phi(x)\,dx,$$

the summation or integration being taken over the range of possible values of $x$. From now on we shall employ another kind of moment, the factorial moment, which has been considered by several writers (11.p.20; 12; 13.§6), and which is specially suited to the case in which the frequencies of the distribution are given for discrete, equidistant values of the variable. The $r^{th}$ factorial moment, for the case where $x$, measured from some arbitrary origin, can increase by increments $h, 2h, \cdots\cdots$, will be defined to be

$$\Sigma\phi(x).\,x(x-h)(x-2h)\cdots(x-\overline{r-1}h),$$

where the summation extends over all possible values of $x$, and will be denoted by $m_{(r)}$. By a suitable choice of scale the increments of $x$ may be taken as equal to unity in any given case.

If $F(t)$ is the frequency generating function of a distribution, that is

$$F(t) = \Sigma\phi(x)t^x,$$

(since we are concerned now only with discontinuous distributions) the factorial moment generating function (referred to afterwards as the factorial m.g.f.) is obtained by the substitution $t = 1+\alpha$. For, making this substitution in the above equation, we have

$$F(1+\alpha) = \sum \phi(x)(1+\alpha)^x$$
$$= \sum \phi(x) + \alpha \sum x\phi(x) + \frac{\alpha^2}{2!}\sum x(x-1)\phi(x) + \cdots,$$

where the coefficient of $\alpha^r/r!$ is $m_{(r)}$ ; that is, we have the factorial m.g.f. for $\phi(x)$.

Evidently the factorial m.g.f. of a sum, or of a linear combination of two or more frequency functions is the same sum or linear combination of the factorial m.g.f.'s of the several frequency functions.

If we denote by $G(f)$ the factorial m.g.f. of $f(x)$, where $f(x)=0$ for $x<0$, then we have the result that

$$G(\nabla^n f) = (-\alpha)^n G(f),$$

where $\nabla f(x) = f(x) - f(x-1)$, that is the 'receding' difference of $f(x)$. This follows at once since

$$G(\nabla^n f) = \sum_x \left\{ f(x) - \binom{n}{1} f(x-1) + \cdots + (-)^n f(x-n)\right\}(1+\alpha)^x$$
$$= \left\{1 - \binom{n}{1}(1+\alpha) + \binom{n}{2}(1+\alpha)^2 - + \cdots + (-)^n(1+\alpha)^n\right\} G(f)$$
$$= (-\alpha)^n G(f).$$

Supposing then that we have obtained the factorial m.g.f. of some frequency distribution as the sum of terms such as $\alpha^r \psi(\alpha)$, and that we know $\psi(\alpha)$ to be the factorial m.g.f. of a certain frequency function $\phi(x)$, we may infer from the above that the frequency function in question is the sum of terms like $(-)^r \Delta^r f(x)$.

If $f(x,y)$ is a function of two variables $x$ and $y$, and $f(x,y)$ has the frequency generating function $\sum_x \sum_y f(x,y) t^x u^y$ by substituting $1+\alpha$ for $t$ and $1+\beta$ for $u$ we shall obtain what may be called a double factorial m.g.f., in which the coefficient of $\alpha^r\beta^s/r!s!$, denoted by $m_{(r,s)}$ - a double factorial moment - will have the form

$$m_{(r,s)} = \sum_x \sum_y x(x-1)\cdots(x-r+1)y(y-1)\cdots(y-s+1) f(x,y).$$

In a later paragraph it will be shown that the factorial moments of a given distribution can be obtained by a simple method of successive summation.

# CHAPTER II.

## THE POISSON EXPONENTIAL FUNCTION.

### §6. THE POISSON LIMIT OF A BINOMIAL.

This limit function was first given by Poisson in his "Recherches sur la Probabilité des Jugements" and was derived to fit the cases where the normal probability function failed to give a close approximation. The assumption made is, that if we consider samples of $N$ drawn from a Bernoullian population where the probability of a success is $p_1$, then $p_1$ is of such an order that $Np_1$ remains finite for large $N$.

The frequency generating function of such samples is $(p_0+p_1t)^N$, where $p_0+p_1=1$, and so it may be written $[1+p_1(t-1)]^N$. Let $Np_1=a$, where $a$ remains finite as $N\to\infty$. Then the frequency generating function can be written $\left[1+\frac{a}{N}(t-1)\right]^N$, which in the limit, as $N\to\infty$, takes the form $e^{a(t-1)}$. The coefficient of $t^x$ in this, that is, the relative frequency of a value $x$ of the independent variable, is $e^{-a}a^x/x!$ which is Poisson's limit function.

This limit can also be derived by considering the limiting form of samples from a population in which the probability of success varies from trial to trial within the sample, but remains constant from sample to sample; the assumption becomes now that $\sum_{r=1}^{N} p_r$, where $p_r$ is the probability of success at any one trial, remains finite as $N$, the number in sample, becomes large. Later this limit function will be obtained under more general considerations.

The factorial m.g.f. of this distribution takes the very simple form $e^{a\alpha}$, so that the successive factorial moments form a geometric progression in $a$, the first factorial moment, or mean, of the distribution. That the second factorial moment is the square of the first, can be used as a criterion of the suitability of fitting any distribution by the Poisson function.

The moment generating function will be $e^{a(e^{\alpha}-1)}$ or $e^{a(\alpha+\alpha^2/2!+\cdots)}$, whence if we refer the distribution to its mean as origin we see that the squared standard deviation, usually denoted by $\sigma^2$, is $a$, that is, equal to the mean of the distribution, as is also the third moment. It also follows at once that the semi-invariants (14) of the distribution are all equal to $a$.

## §7. NUMERICAL EXAMPLE.

The following data were obtained by an experiment approximating to the conditions for the Poisson function. Two hundred marbles were taken, of which five were distinguished from the remainder in colour. These were thoroughly mixed in a bag and poured out into grooves on a board capable of containing rows of twenty marbles each. The number of specially coloured marbles in each row was noted. This was repeated twenty times giving in all two hundred observations. A Poisson function is to be fitted to the results.

| (1) $x$ | (2) $f(x)$ | (3) | (4) | (5) |
|---|---|---|---|---|
| 0 | 120 | 200 | | |
| 1 | 63 | 80 | 100 | |
| 2 | 15 | 17 | 20 | 24 |
| 3 | 1 | 2 | 3 | 4 |
| 4 | 1 | 1 | 1 | 1 |

The data are in column (2). Column (3) is obtained from (2) by successive summation from the bottom. The total is clearly $N = \Sigma f(x)$. Column (4) is obtained from (3) similarly, but the process is stopped one line below the top figure in the previous column. This gives $\Sigma x f(x)$, that is, $N m_{(1)}$. Column (5) is obtained in the same way, but the process is stopped one line lower than the top of the previous column. This gives $\Sigma x(x-1)f(x)/2!$, that is $N m_{(2)}/2!$. The procedure is quite general, and so factorial moments of higher orders can be obtained in the form $N m_{(r)}/r!$.

Thus here we have,

$$m_{(1)} = \frac{100}{200} = \cdot 5, \qquad m_{(2)} = \frac{24}{200} \times 2! = \cdot 24.$$

The values of $e^{-a} a^x/x!$ can be obtained from tables, such as H.E.Soper's in "Biometrika" X, where it is tabulated for $a = 0{\cdot}1, 0{\cdot}2, \ldots, 15{\cdot}0$ and for all integral $x$ giving a figure in the sixth decimal place. In this example $a = \cdot 5$. The remaining arithmetical work in fitting the curve is given in the following table:

| $x$ | (1) $e^{-a}a^x/x!$ | $200 \times (1)$ | Data. |
|---|---|---|---|
| 0 | ·606 | 121 | 120 |
| 1 | ·303 | 61 | 63 |
| 2 | ·076 | 15 | 15 |
| 3 | ·013 | 3 | 1 |
| 4 | ·002 | 0 | 1 |

A comparison with the observed values in the last column shows the fit to be quite as good as could be expected.

§8. THE MEDIAN AND THE MODE.

Although the Poisson limit function is usually considered only for discrete values of the argument, it is of interest to consider it from the point of view of continuous distributions and to establish approximate results for the position of the median and the mode.

Let us take first of all the theoretical Poisson distribution of a sample of 1000 from a population with mean equal to 4.5. This distribution is

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\psi(x)$ | 11 | 50 | 113 | 169 | 190 | 171 | 128 | 82 | 46 | 23 | 10 | 4 | 2 | 1 | 1000 |

The corresponding cumulative distribution is

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\Psi(x)$ | 11 | 61 | 174 | 343 | 533 | 704 | 832 | 914 | 960 | 983 | 993 | 997 | 999 | 1000 |

In order to find the median and the mode we shall have to consider that the data represent frequencies grouped about the given values of the argument, so that the abscissa of each value of the cumulative table is really .5 ahead of the abscissa of the corresponding class. If we apply the Bessel interpolation formula to values in the following table, taken from the cumulative distribution, with origin at $x = 4$,

| $x$ | $\Psi(x)$ | $\Delta$ | $\Delta^2$ |
|---|---|---|---|
| 3.5 | 343 | | 21 |
| ← | (438) | 190 | (1) |
| 4.5 | 533 | | −19 |
| | | 171 | |
| 5.5 | 704 | | |

then for the median we have,

$$500 = 438 + 190x + \frac{x^2 - 1/4}{2!}(1),$$

so that the median is at $x = 4\frac{1}{3}$ approximately. A similar interpolation using the frequencies in the classes $x = 3, 4, 5$ gives as the approximate position of the mode, $x = 4$. Thus the distances between the mean and median, and the median and mode are approximately in the ratio $1:2$, a result which may be compared with the similar one for distributions not deviating greatly from the normal distribution.

Considered generally, the result that the median is approximately given by $(mean - 1/6)$ can be derived from a very interesting result due to Ramanujan (15), that

$$1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^x}{x!}\theta = \frac{1}{2}e^x,$$

where $\theta = \frac{1}{3} + \frac{4}{135(x+k)}$ and $k$ lies between $8/45$ and $2/21$, depending on the magnitude of $x$. Thus, in the case of Poisson frequencies, for a distribution with mean $a$, we have approximately, neglecting the correction involving $k$,

$$e^{-a}\left(1 + \frac{a}{1!} + \frac{a^2}{2!} + \cdots + \frac{1}{3}\frac{a^a}{a!}\right) = \frac{1}{2}, \text{ (approx.)},$$

a result which gives the approximate position of the median. If the whole of the last term $a^a/a!$ were taken, the median would be at $a + \cdot 5$ but since it is only $1/3$ of this term, we have by proportional parts, that the abscissa of the median is approximately $a + 1/2 - 2/3$, or $a - 1/6$.

The following table shows the corresponding values of

$\frac{1}{2}e^x$ and of the approximation $1+\frac{x}{1!}+\frac{x^2}{2!}+\cdots+\frac{1}{3}\frac{x^x}{x!}$, for $x=1,2,\cdots,10$. It will be seen that the absolute magnitude of the error increases, but relatively it decreases.

| $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\frac{1}{2}e^x$ | 1·36 | 3·69 | 10·05 | 27·3 | 74·2 | 201·6 | 548·5 | 1490·5 | 4051·5 | 11013 |
| Approx. | 1·33 | 3·67 | 10·0 | 27·2 | 74·1 | 201·4 | 547·5 | 1489·0 | 4048·1 | 11005·1 |

In order to obtain a general result for the mode, it will be necessary to find the maximum value of $e^{-a}a^x/\Gamma(x+1)$, substituting $\Gamma(x+1)$ for $x!$ since now we are considering $x$ as varying continuously. Differentiating logarithmically, we see that the mode is given by the value of $x$ which makes

$$\log a - \frac{d}{dx}\log\Gamma(x+1)$$

zero. We now make use of the formula (13.p.61) for $\frac{d}{dx}f(x)$ in terms of the central differences of $f(x)$, namely

$$\frac{d}{dx}f(x+1) = \left(\delta - \frac{1}{24}\delta^3 + \frac{3}{640}\delta^5 - \cdots\right)f(x+1),$$

which has the remainder term

$$(-)^r\,\frac{(1.3.\cdots.\overline{2r-1}.)^2}{2^{2r}(2r+1)!}\,f^{(2r+1)}(\xi),$$

where $\xi$ lies between the greatest and least values of $x$ used in forming the difference table to get as far as $\delta^{2r-1}f$.

If now we put $\log\Gamma(x+1)$ for $f(x+1)$, and take only the first term of the above formula, we have, since

$$\delta\log\Gamma(x+1) = \log\Gamma(x+3/2) - \log\Gamma(x+1/2) = \log(x+1/2),$$

$$\frac{d}{dx}\log\Gamma(x+1) = \log(x+1/2),$$

with a remainder term $-\frac{1}{24}\frac{d^3}{dx^3}\log\Gamma(\xi)$, where $x<\xi<x+1$, which even for moderate $x$ soon becomes small. Hence, finally, we have for the abscissa of the mode $x = a - 1/2$. Combining these, we have the general result for Poisson distributions that the distances between the mean and median and the median and mode are in the approximate ratio of $1:2$.

## §9. THE POISSON LIMIT OF A MULTINOMIAL.

The form of the Poisson function given in §6 was derived by considering the limiting form of samples of $N$ from what were essentially populations of a binomial nature. If we now consider populations where the character involved can assume a greater range of values than merely 0 and 1, that is, populations of a multinomial character, where the probabilities of non-zero values are of the small order postulated for the Poisson limit, we obtain quite a different form of the limit function.

Let us consider first the case where the character may assume the values $0, 1, 2, \ldots l$, with probabilities $p_0, p_1, \ldots, p_l$, then the frequency generating function of samples of $N$ from such a population is

$$(p_0 + p_1 t + p_2 t^2 + \cdots + p_l t^l)^N,$$

where $p_0 + p_1 + p_2 + \cdots + p_l = 1$, and $p_1, p_2, \cdots p_l$, are of such an order that $Np_1, Np_2, \cdots Np_l$ all remain finite, equal to $a_1, a_2, \cdots, a_l$ say, when $N$ becomes large, so that the frequency generating function will in the limit have the form

$$e^{a_1(t-1) + a_2(t^2-1) + \cdots + a_l(t^l-1)}.$$

The coefficient of $t^x$ in this will give the relative frequency of a value $x$ of the argument, in any type of population of this kind that one wishes to consider. In the case of a trinomial population, the frequency function is

$$e^{-(a_1+a_2)}\left(\frac{a_1^{x}}{x!} + \frac{a_1^{x-2}}{(x-2)!}\cdot\frac{a_2}{1!} + \frac{a_1^{x-4}}{(x-4)!}\cdot\frac{a_2^2}{2!} + \cdots\right).$$

If, instead of as above, we suppose that the variable can assume the values $0, 1/N, 2/N, \cdots, l/N$ we have as frequency generating function of samples of $N$ from such a population

$$\left(p_0 + p_1 t^{1/N} + p_2 t^{2/N} + \cdots + p_l t^{l/N}\right)^N,$$

or, $$\left[1 + p_1(t^{1/N}-1) + \cdots + p_l(t^{l/N}-1)\right]^N.$$

This will have the factorial m.g.f.

$$\Big[1 + p_1\left\{\alpha/N + \tfrac{\alpha^2}{2!}\, 1/N(1/N-1) + \cdots\right\} + \cdots$$
$$+\; p_l\left\{\alpha\cdot l/N + \tfrac{\alpha^2}{2!}\, l/N(l/N-1) + \cdots\right\}\Big]^N.$$

Now on the assumption that $\sum_{r=1}^{l} p_r\cdot r/N$ is of order $1/N$, equal to $m_{(1)}/N$ say, so $\sum_{r=1}^{l} p_r\cdot r/N(r/N-1)$ is of order $1/N$, equal to $m_{(2)}/N$ say, and so on, we have as the limiting form of the factorial m.g.f.

$$e^{m_{(1)}\alpha + m_{(2)}\alpha^2/2! + \cdots}.$$

In the particular case where $m_{(1)} = m_{(2)} = \cdots = m$, that is the factorial moments of the probabilities are all equal, the factorial m.g.f. takes the form

$$e^{m(e^{\alpha}-1)} = e^{-m}\, e^{me^{-1}e^{\alpha+1}},$$

whence by equating this to the general form of the factorial m.g.f. and solving, we obtain the frequency function

$$\phi(x) = e^{-m}\left[\frac{me^{-1}}{1!}\cdot\frac{1}{x!} + \frac{m^2e^{-2}}{2!}\cdot\frac{2^x}{x!} + \cdots + \frac{m^r e^{-r}}{r!}\cdot\frac{r^x}{x!} + \cdots\right]$$

$$= \frac{e^{-(m+1)}}{x!}\cdot m\left[1 + \frac{me^{-1}}{1!}\cdot 2^{x-1} + \frac{m^2e^{-2}}{2!}\, 3^{x-1} + \cdots + \frac{m^{r-1}e^{-\overline{r-1}}}{(r-1)!}\, r^{x-1} + \cdots\right].$$

Thus it is seen that the Poisson ~~Unit~~ limit function established in §6 is not based on the same generality of assumptions as the normal function. As usually derived (16) the Poisson ~~Unit~~ limit function gives the distribution of a variable based on the assumption that it is compounded of a large number of elementary increments, each following a binomial distribution, that is, having the possible values $0$ or $1$, and with the probability of obtaining the value unity, of the small order postulated above.

## §10. THE DISTRIBUTION AND FACTORIAL MOMENTS OF $m_{(1)}$ IN SAMPLES.

In this paragraph we shall investigate the distribution and the factorial m.g.f. of the first factorial moment $m_{(1)}$, as computed from samples of $N$ drawn from an infinite Poisson universe.

The universe or population being typified by a frequency generating function $e^{a(t-1)}$, the frequency generating function of samples of $N$ is, by compound probability,

$$\left\{e^{a(t-1)}\right\}^N \text{, or } e^{Na(t-1)} \quad \text{- - - - - - - - -} \quad (10\cdot1).$$

This is the distribution of $(x_1+x_2+\cdots+x_N)$, or $N m_{(1)}$. Hence the distribution of $m_{(1)}$ is obtained by altering the scale of measurement in (10·1) in the ratio $N:1$, that is, we must replace $t$ by $t^{1/N}$. Thus the distribution of $m_{(1)}$ is given by

$$e^{Na(t^{1/N}-1)} \quad \text{- - - - - - - - - - - -} \quad (10\cdot2).$$

From this we learn at once that the relative frequency, or probability, of a value $x$ for $m_{(1)}$ is given by

$$e^{-Na}(Na)^{Nx}/(Nx)!\,.$$

Further by substituting $1+\beta$ for $t$ in (10·2) we derive the factorial m.g.f. of $m_{(1)}$ in samples of $N$ as

$$e^{-Na}\,e^{Na(1+\beta)^{1/N}}.$$

The first factorial moment of this sample distribution, which we may denote by $\underline{m_{(1)}}$, is the coefficient of $\beta$ in the expansion of this and is seen to be equal to $a$. Thus the expected, or mean value of the mean in samples of $N$, is no different from the mean of the infinite Poisson universe itself.

The second factorial moment of the sample distribution differs from that of the universe, for the coefficient of $\beta^2/2!$ in the generating function is

$$e^{-Na}\left\{Na.\tfrac{1}{N}\left(\tfrac{1}{N}-1\right)+\tfrac{N^2a^2}{2!}\cdot\tfrac{2}{N}\left(\tfrac{2}{N}-1\right)+\cdots\right\}$$

$$=ae^{-Na}\left\{1/N-1+\tfrac{Na}{1!}(2/N-1)+\tfrac{N^2a^2}{2!}(3/N-1)+\cdots\right\}$$

$$=ae^{-Na}\left\{1/N.e^{Na}(Na+1)-e^{Na}\right\}$$

$$=a(a-1)+a/N.$$

For samples of one, that is single drawings, this gives the familiar result $m_{(2)} = a^2$.

In the same way the following results for the third and fourth factorial moments respectively, of this distribution may be obtained,

$$a(a-1)(a-2)\ +3a(a-1)/N\ +a/N^2,$$

and

$$a(a-1)(a-2)(a-3)+6a(a-1)(a-2)/N+7a(a-1)/N^2+a\ 1/N(1/N-1),$$

which reduce to $a^3$ and $a^4$ respectively when $N=1$, as we should expect. Though the coefficients (1,1), 1,3,1), and (1,6,7,1) are familiar as the <u>divided</u> "differences of zero", the general formula does not appear to be simple.

# CHAPTER III.

## THE CHARLIER SERIES OF TYPE B.

### §11. DERIVATION OF THE TYPE B SERIES.

The dominant part of the Charlier series of Type B is the Poisson limit function established in §6; but both the Poisson function and the Charlier series arise under more general conditions.

Let the variable $x$ be the resultant of $N$ fortuitous increments $\delta x$, where $\delta x_j$, the $j^{th}$ of these increments, may take any one of $k$ values, each with a certain probability. The assumptions are, that the elementary increments $\delta x_j$ may take the values $0, 1/N, 2/N, \ldots$ ; also, that the probability of non-zero increments is of so small an order in $N$, that if $\delta m_{(j:r)}$ is the factorial moment of order $r$ of the elementary distribution, then

$$\delta m_{(j:1)} = \sum_{1}^{k} \delta x_j \, p_j \qquad \text{is of order } 1/N,$$

$$\delta m_{j:2} = \sum_{1}^{k} \delta x_j (\delta x_j - 1/N) \, p_j \qquad \text{is of order } 1/N^2 \text{, etc.}$$

For reasons of convenience that will appear later we write the elementary factorial m.g.f., namely,

$$F_j(\alpha) = 1 + \delta m_{(j:1)} \alpha + \delta m_{(j:2)} \alpha^2/2! + \cdots$$

as

$$F_j(\alpha) = e^{\delta m_{(j:1)} \alpha} (1 + b_2 \alpha^2 + b_3 \alpha^3 + \cdots),$$

where $b_2$ is of order $1/N^2$, $b_3$ of order $1/N^3$, and so on.

By compound probability the factorial m.g.f. of a variable which is the sum of independent variables, is the product of the factorial m.g.f.'s of the several variables. Hence the factorial m.g.f. of $x$ in the present case is

$$F(\alpha) = e^{\sum_{\nu=1}^{q} \delta m_{(\nu:1)}\alpha} \prod (1+b_2\alpha^2+b_3\alpha^3+\cdots)$$

$$= e^{a\alpha}(1+B_2\alpha^2+B_3\alpha^3+\cdots),$$

where $a$ is the mean value of $x$, given by $a = \sum \delta m_{(\nu:1)}$.

Reserving consideration of the question of the order in $N$ of $B_2, B_3, \cdots$, we shall find what frequency function corresponds to the factorial m.g.f. $F(\alpha)$ as given above. When $N$ is large, the dominant term, it will appear, is the first term $e^{a\alpha}$. If $\psi(x)$ is the frequency function corresponding to this, we shall have

$$e^{a\alpha} = \sum_{x=0}^{\infty} \psi(x)(1+\alpha)^x,$$

i.e. $$e^{-a}\, e^{a(1+\alpha)} = \sum_{x=0}^{\infty} \psi(x)(1+\alpha)^x,$$

whence $$\psi(x) = e^{-a}a^x/x!,$$

which is Poisson's function given earlier. This is the usual form in which it is given. We ought, however, to take account here of the fact that $x$, being compounded of elements which can assume the values $1/N, 2/N, \cdots$, can also assume fractional values, and so we adopt the more general form,

$$\psi(x) = e^{-a}a^x/\Gamma(x+1) \quad (x>0).$$

In order to obtain the more general frequency function which results from taking all the terms of $F(\alpha)$, use is made of the result given in §5. This yields the more general frequency function,

$$\phi(x) = \psi(x) + B_2 \Delta^2 \psi(x) - B_3 \Delta^3 \psi(x) + \cdots,$$

where $\psi(x)$ is the Poisson function found above.

It remains to investigate the order in $N$ of the coefficients $B_r$, where

$$1 + B_2\alpha^2 + B_3\alpha^3 + \cdots = (1 + b_2\alpha^2 + b_3\alpha^3 + \cdots)^N.$$

Clearly $B_r$ will be composed of terms like $b_h b_k \cdots b_q$, where $h, k, \cdots q$ may take the integer values $2, 3, 4, \cdots$, and where the sum of these indices $h, k, \cdots q$ is $r$. The order of each such term is evidently $1/N^r$. Also, if the number of parts in such a partition of $r$ is $m$, the number of $m$-part terms is ${}^NC_m$, which is of order $N^m$. This order is greatest when $m$ is greatest, in other words, the order of $B_r$ is determined by the partition of the integer $r$ which has most parts. If $r$ is even, equal to $2s$, such a partition is made up of $s$ $2$'s; if $r$ is odd of the form $2s-1$, the partition is of $s-2$ $2$'s and a single $3$, that is $s-1$ parts. It follows at once that the order both of $B_{2s-1}$ and of $B_{2s}$ is $1/N^s$, so that the $B$'s descend in order regularly in pairs (10), $B_2$ being of order $1/N$, $B_3$ and $B_4$ of order $1/N^2$, and so on. This is why in fitting a curve of Type B it is advisable to stop after an even coefficient.

## §12. RELATIONS BETWEEN THE COEFFICIENTS AND FACTORIAL MOMENTS IN TYPE B.

For the purpose of fitting a series of Type B to a given distribution it is necessary to know the relations (17) between the coefficients $B_r$ and the factorial moments as computed from the distribution.

These are found by considering the factorial m.g.f. found above. If in this we expand the exponential, remembering that $m_{(r)}$ is the coefficient of $\alpha^r/r!$, we derive the relations

$$m_{(1)} = a,$$
$$m_{(2)} = a^2 \quad +2!B_2,$$
$$m_{(3)} = a^3 \quad +\tfrac{3!}{1!}aB_2 + 3!B_3,$$
$$m_{(4)} = a^4 \quad +\tfrac{4!}{2!}a^2B_2 + \tfrac{4!}{1!}aB_3 + 4!B_4,$$

and generally,

$$m_{(r)} = a^r \quad + \frac{r!}{(r-2)!}a^{r-2}B_2 + \cdots\cdots + \frac{r!}{1!}aB_{r-1} + r!B_r.$$

These expressions may be made formally complete by inserting after the first term in each case $B_1$, $\frac{2!}{1!}aB_1$, $\frac{3!}{2!}a^2B_1$, ......, $\frac{r!}{(r-1)!}a^{r-1}B_1$, respectively, where $B_1$ is zero. The last of these relations may then be written

$$m_{(r)}/a^r = 1 + \binom{r}{1}\frac{1!B_1}{a} + \binom{r}{2}\frac{2!B_2}{a^2} + \cdots + \binom{r}{1}\frac{(r-1)!B_{r-1}}{a^{r-1}} + \frac{r!B_r}{a^r}.$$

If in this we take $m_{(s)}/a^s$ and $s!B_s/a^s$ as the variables, then by the application of a reciprocal result in matrices having binomial coefficients for elements, we may invert the relations, and so obtain

$$r!B_r/a^r = (-)^r\left\{1 - \binom{r}{1}\frac{m_{(1)}}{a} + \binom{r}{2}\frac{m_{(2)}}{a^2} - + \cdots + (-)^{r-1}\binom{r}{1}\frac{m_{(r-1)}}{a^{r-1}} + (-)^r\frac{m_{(r)}}{a^r}\right\},$$

or, $\quad r!\,B_r = m_{(r)} - \binom{r}{1}m_{(r-1)}.a + \cdots + (-)^{r-1}\binom{r}{1}m_{(1)}a^{r-1} + (-)^r a^r.$

Since $m_{(1)} = a$, the last two terms may be combined into $(-)^{r-1}(r-1)a^r$. The expression for $B_r$ in terms of the factorial moments can be represented symbolically by

$$B_r = \frac{([m]-a)^r}{r!} \quad \cdots\cdots\cdots\cdots \quad (12.1)$$

where $[m]^s$ is to be understood as $m_{(s)}$.

## § 13. THE TYPE B. POLYNOMIALS.

Let us expand formally in powers of $\alpha$, $G(t)e^{-a\alpha}$, where $G(t) = \sum_x f(x)(1+\alpha)^x$ is the factorial m.g.f. of the function $f(x)$.

Thus

$$G(t)e^{-a\alpha} = \left\{\sum_{x=0}^{\infty} f(x)(1+\alpha)^x\right\}\left\{\sum_{r=0}^{\infty}(-)^r a^r\alpha^r/r!\right\}$$

$$= \left\{\sum_{n=0}^{\infty}\frac{\alpha^n}{n!}\left(\sum_{x=0}^{\infty} f(x).x^{[n]}\right)\right\}\left\{\sum_{r=0}^{\infty}(-)^r a^r\alpha^r/r!\right\}$$

$$= \sum_{n=0}^{\infty} B_n\alpha^n \qquad (13\cdot 1),$$

where

$$B_n = \sum_{x=0}^{\infty} f(x)\left(\frac{x^{[n]}}{n!} - \frac{x^{[n-1]}}{(n-1)!}\cdot\frac{a}{1!} + \frac{x^{[n-2]}}{(n-2)!}\cdot\frac{a^2}{2!} - \cdots + \cdots + (-)^n\frac{a^n}{n!}\right)$$

$$= \frac{a^n}{n!}\sum_{x=0}^{\infty} f(x).\,K_n(x) \qquad (13\cdot 2),$$

in which $x^{[r]} = x(x-1)\cdots(x-r+1)$,

and $K_n(x) = \left(x^{[n]} - \binom{n}{1}x^{[n-1]}.a + \binom{n}{2}x^{[n-2]}.a^2 + \cdots + (-)^n a^n\right)/a^n$

is a polynomial in $x$ of degree $n$.

Now writing (13·1) in the form

$$G(t) = \sum_{n=0}^{\infty} B_n\alpha^n e^{a\alpha},$$

and remembering the result found for factorial m.g.f. in § 5, we thus obtain the formal expansion for $f(x)$,

$$f(x) = \psi(x) - B_1\Delta\psi(x) + B_2\Delta^2\psi(x) + \cdots \qquad (13\cdot 3).$$

where $\psi(x) = e^{-a}a^x/x!$ and the coefficient $B_n$ is given by (13·2).

A sufficient condition for the uniform convergence of this expansion has been given (4.p.110) and is that

$$\sqrt[n]{m_n} < \frac{a}{e}\log n' \text{ where } n'\log n' = n \quad ,$$

and $m_n = \int_0^\infty x^n f(x)\,dx$, that is, $m_n$ is the $n^{th}$ moment of $f(x)$.

The expansion thus obtained is the Charlier Type B. expansion for a given function, $f(x)$ , in terms of the successive receding differences of the Poisson exponential function. The polynomials, $K_n(x)$, which arise in this expansion have a part in the Type B. series (17, 18, 19) quite analogous to that of the Hermite polynomials in the Type A. series.

From (13·2) and (13·3) we have the orthogonality relation that these polynomials fulfil, namely,

$$\sum_x K_r(x) \Delta^s \psi(x) = 0 \qquad r \neq s,$$
$$= r!/a^r \qquad r = s.$$

By the nature of their formation, or by a direct application of results of the finite difference calculus, one can easily show that

$$\Delta^s \psi(x) = (-)^s K_s(x) \cdot \psi(x),$$

which is the customary definition of the Type B. polynomial of order $s$ . The orthogonality condition may now be written

$$\sum_x K_r(x) K_s(x) \psi(x) = 0 \qquad r \neq s,$$
$$= (-)^r r!/a^r \qquad r = s.$$

Further properties of the Type B. polynomials are given below.

From the definition of the polynomials we have

$$\Delta K_r(x) = \frac{r}{a} K_{r-1}(x-1)$$

or employing advancing differences,

$$\triangleright K_r(x) = \frac{r}{a} K_{r-1}(x) \text{..........} \qquad (13 \cdot 4)$$

a result which is useful in computing the values of the polynomials.

A further result that provides a partial check in computing these values is,

$$K_r(r+1) = -K_{r+1}(r).$$

This follows at once when the right hand side is written out in full and we use the fact that $\binom{r+1}{s+1} r^{[r-s]} = \binom{r}{s}(r+1)^{[r-s]}$.

$$\sum_x K_r(x)\,\psi(x) = \sum_x \left\{ \psi(x) - \binom{r}{1}\psi(x-1) + \cdots + (-)^r \psi(x-r) \right\}$$
$$= (1-1)^r$$
$$= 0,$$

from the definition of the polynomials and using $\sum_x \psi(x) = 1$.

The polynomials have the generating function

$$e^{-t}\left(1+\frac{t}{a}\right)^x = \sum_r K_r(x)\, t^r/r!, \qquad (13\cdot5)$$

and can be represented by the operational equation.

$$K_r(x) = e^{-a\Delta} x^{[r]}/a^r. \qquad (13\cdot6)$$

The results (13·5) and (13·6) may be compared with those corresponding for the Hermite polynomials which have the generating function $e^{-\frac{1}{2}t^2}.e^{tx}$ and the operational form $H_r(x) = e^{-\frac{1}{2}D^2}.x^r$, where $D \equiv \frac{d}{dx}$. By such a comparison the appropriateness of the difference calculus and factorials for the Type B. series, and that of the differential calculus and powers for the Type A, is evident.

Using (13·5) we can obtain an expression connecting the polynomials when $x$ is referred to a new origin with those referred to the old origin. If the new origin is at $x=\alpha$ then the generating function of the polynomials becomes

$$e^{-t}\left(1+\frac{t}{a}\right)^{x-\alpha} \quad \text{or} \quad e^{-t}\left(1+\frac{t}{a}\right)^x\left(1+\frac{t}{a}\right)^{-\alpha},$$

whence

$$K_r(x-\alpha) = K_r(x) - \frac{\alpha.r}{1!a}K_{r-1}(x) + \frac{\alpha(\alpha+1)r(r-1)}{2!\,a^2}K_{r-2}(x) - \cdots;$$

as might also be proved by combining the Gregory-Newton interpolation formula with the result (13·4).

By inverting the operational result (13·6), together with (13·4), we have an expression for the factorial $x^{[r]}$ in terms of the Type B. polynomials,

$$x^{[r]} = a^r\left(K_r(x) + \binom{r}{1} a K_{r-1}(x) + \cdots + a^r\right),$$

which may be compared with the corresponding result for $x^r$ in terms of Hermite polynomials, obtained similarly by inverting the operational expression for $H_r(x)$ given above,

$$x^r = H_r(x) + \binom{r}{2}\cdot\frac{1}{2}\cdot H_{r-2}(x) + \binom{r}{4}\frac{1}{2^2}\cdot H_{r-4}(x) + \cdots .$$

If we differentiate (13·5) with respect to $t$ then we have

$$-e^{-t}\left(1+\frac{t}{a}\right)^x + \frac{x}{a}e^{-t}\left(1+\frac{t}{a}\right)^{x-1} = \sum_n K_n(x)\, t^{n-1}/(n-1)!,$$

whence we obtain the recurrence relation

$$-K_n(x) + \frac{x}{a}K_n(x-1) = K_{n+1}(x),$$

i.e.

$$K_{n+1}(x) + K_n(x) - \frac{x}{a}K_n(x-1) = 0.$$

But by (13·4)

$$K_n(x) - K_n(x-1) = \frac{n}{a}K_{n-1}(x-1),$$

and making this substitution in the above formula we obtain an alternative form for the recurrence relation

$$K_{n+1}(x) - K_1(x)K_n(x-1) + \frac{n}{a}K_{n-1}(x-1) = 0.$$

Now (13·2) the expression for $B_n$, the coefficient in the expansion of the function with factorial m.g.f. $G(t)$ may be written

$$B_n = \frac{m_{(n)}}{n!} - \frac{m_{(n-1)}}{(n-1)!}\cdot\frac{a}{1!} + \cdots + (-)^n\frac{a^n}{n!},$$

where $m_{(r)}$ is the $r^{th}$ factorial moment of the function $f(x)$. In §11 we obtained the factorial m.g.f. of the generalized Poisson function as $e^{a\alpha}(1+B_2\alpha^2+B_3\alpha^3+\cdots)$ so that substituting this for $G(t)$ we have at once, that the coefficients are connected with the factorial moments of the frequency distribution, by the relation,

$$B_r = \frac{m_{(r)}}{r!} - \frac{m_{(r-1)}}{(r-1)!}\frac{a}{1!} + \cdots + (-)^r \frac{a^r}{r!},$$

the result that was established in §12.

## §14. THE TYPE B. SERIES AND THE LEXIS DISTRIBUTION.

Consider samples of $s$ each drawn from $N$ different Bernoullian populations of such a kind that the probability of obtaining the value unity in the $r^{th}$ sample, $p_r$ say, is small, of such an order that $s p_r$ remains finite for large values of $s$. It is desired to find the effect that the varying probability has on the moments of the resulting distribution.

The relative frequency of a value $x$, $\phi_r(x)$, from the $r^{th}$ of these populations will be given by the Type B. series,

so that $\quad \phi_r(x) = \psi_r(x) + B_2^{(r)} \Delta^2 \psi_r(x) + \cdots$

where $\quad \psi_r(x) = e^{-a_r} a_r^{\,x}/x!$, and $a_r = s p_r$.

This will have the factorial m.g.f.

$$(1+p_r\alpha)^s \quad \text{or} \quad e^{a_r\alpha}\left(1 + B_2^{(r)}\alpha^2 + \cdots\right),$$

whence
$$m_{(2)}^{(r)} = s(s-1)\,p_r^2$$
$$= a_r^2 - a_r p_r$$
$$= 2B_2^{(r)} + a_r^2.$$

If $f(x)$ is the frequency distribution of these samples, then as a value $x$ is equally likely to be obtained from any one of them $\quad f(x) = \frac{1}{N}\left(\phi_1(x) + \phi_2(x) + \cdots + \phi_N(x)\right)$

so that $f(x)$ will have the factorial m.g.f.

$$\frac{1}{N}\left\{\sum_{r=1}^{N} e^{a_r\alpha}\left(1 + B_2^{(r)}\alpha^2 + \cdots\right)\right\}.$$

$f(x)$ being represented by a Type B. series, it will have a factorial m.g.f. of the form

$$e^{a\alpha}(1+B_2\alpha^2+\cdots).$$

A comparison of the two forms of the factorial m.g.f. shows that the first factorial moment of $f(x)$,

$$m_{(1)} = a = \frac{1}{N}(a_1+a_2+\cdots+a_N) = \frac{s}{N}(p_1+p_2+\cdots+p_N) = sp,$$

where $p$ is the average of the varying probabilities. Similarly the second factorial moment

$$m_{(2)} = \frac{1}{N}\sum_{r=1}^{N}(2B_2^{(r)}+a_r^2)$$
$$= \frac{1}{N}\sum_{r=1}^{N} s(s-1)p_r^2.$$

Assuming that the probability remains constant from sample to sample, with the value $p$, that is, that we are sampling from a Bernoullian population, then we would have

$$m_{(2)B} = s(s-1)p^2.$$

The identity $\sum_{r=1}^{N} p_r^2 = Np^2 + \sum_{r=1}^{N}(p_r-p)^2$

gives at once that

$$m_{(2)} = m_{(2)B} + \frac{s(s-1)}{N}\sum_r (p_r-p)^2.$$

These results are of a similar character to those found in sampling from an ordinary Lexis population, namely, that the first factorial moment of the samples of the Lexian population is the same as that of the Bernoullian population in which the probability of a non-zero value is taken as the average of the varying probabilities of the Lexis population; while the second factorial moment exceeds that of the corresponding Bernoullian series, by an amount proportional to the sum of the squared deviations from their mean value of the varying probabilities.

## §15. NUMERICAL PROCESS OF FITTING A SERIES OF TYPE B.

A curve of Type B. is to be fitted to data given in W. Palin Elderton's "Frequency Curves" and "Correlation" (2nd Ed). p.131. We shall take only one correction term, that involving $B_2$. The factorial moments are calculated from the following table:

| $x$ | $f(x)$ | | | |
|---|---|---|---|---|
| 0 | 133 | 222 | | |
| 1 | 55 | 89 | 140 | |
| 2 | 23 | 34 | 51 | 76 |
| 3 | 7 | 11 | 17 | 25 |
| 4 | 2 | 4 | 6 | 8 |
| 5 | 2 | 2 | 2 | 2 |

Thus

$$m_{(1)} = 140/222 = 0.631, \quad m_{(2)} = 152/222 = 0.684,$$

and so $$B_2 = \tfrac{1}{2}\left(.684 - (.631)^2\right) = 0.143.$$

The remaining arithmetical work in fitting the curve is given in the following table:

| $x$ | (1) $\psi(x)$ | (2) $(1)\times N$ | (3) $\Delta(2)$ | (4) $\Delta^2(2)$ | (5) $B_2\times(4)$ | (6) $(2)+(5)$ | (7) Data. |
|---|---|---|---|---|---|---|---|
| 0 | ·5320 | 118·1 | 118·1 | 118·1 | 16·8 | 134·9 | 135 |
| 1 | ·3357 | 74·5 | -43·6 | -161·7 | -23·1 | 51·4 | 55 |
| 2 | ·1059 | 23·5 | -51·0 | -7·4 | -1·0 | 22·5 | 23 |
| 3 | ·0223 | 4·9 | -18·6 | 32·4 | 4·6 | 9·5 | 7 |
| 4 | ·0035 | ·8 | -4·1 | 14·5 | 2·1 | 2·9 | 2 |
| 5 | ·0004 | ·1 | -·7 | 3·4 | ·5 | ·6 | 2 |

The values of $\psi(x) = e^{-a}a^x/x!$ are obtained from tables with $a = \cdot631$.

Columns (3) and (4) are formed by considering there to be zeros above in column (2) corresponding to negative values of $x$ in $\psi(x)$.

Column (6) gives the graduated figures as obtained using one correction term. A comparison with the original figures, given in (7), shows that they form quite a reasonable fit.

# CHAPTER IV.

## THE TYPE B. CORRELATION FUNCTION.

### §16. THE POISSON CORRELATION FUNCTION.

Let the following fourfold table represent the frequency distribution of a population involving two characters $A$ and $B$ which are not independent,

| | | $A$ | |
|---|---|---|---|
| | $a$ | $b$ | $\bar{p}_2$ |
| $B$ | $c$ | $d$ | $p_2$ |
| | $\bar{p}_1$ | $p_1$ | $1$ |

so that $p_1$ and $p_2$ represent the relative frequency of $A$ and $B$ respectively, and $\bar{p}_1, \bar{p}_2$ that of the absence of the characters. Owing to the existence of some degree of dependence between the characters, the relative frequency $d$ will be different from $p_1p_2$. The frequency generating function of the population will be

$$a + bA + cB + dAB,$$

or

$$a + (p_1 - d)A + (p_2 - d)B + dAB,$$

or

$$1 + p_1(A-1) + p_2(B-1) + d(A-1)(B-1).$$

Hence the factorial m.g.f. of the population is

$$1 + p_1\alpha + p_2\beta + d\alpha\beta$$

and so that of samples of $N$ is

$$(1 + p_1\alpha + p_2\beta + d\alpha\beta)^N$$
$$= 1 + Np_1\alpha + Np_2\beta + N(d - p_1p_2)\alpha\beta + (Np_1\alpha + Np_2\beta)^2/2! + Np_1 \cdot N(d-p_1p_2)\alpha^2\beta - Np_1^2\alpha^2/2! - Np_2^2\beta^2/2! + \cdots .$$

Let $Np_1 = m_1$, $Np_2 = m_2$ and $N(d - p_1p_2) = \bar{m}$, then the assumption made as regards the quantities $p_1, p_2$ and $d$ is that they are of such an order that $m_1, m_2$ and $\bar{m}$ remain finite for large values of $N$. Under these conditions the factorial m.g.f. may be written as

$$e^{m_1\alpha + m_2\beta + \bar{m}\alpha\beta}\left(1 + \text{terms of order } 1/N \text{ and higher}\right),$$

so that in the limiting case it takes the form

$$e^{m_1\alpha + m_2\beta + \bar{m}\alpha\beta} \quad \ldots\ldots\ldots\ldots\ldots (16\cdot1).$$

In order to find the frequency function involved we have to solve the equation,

$$e^{m_1\alpha + m_2\beta + \bar{m}\alpha\beta} = \sum_x \sum_y \phi(x,y)(1+\alpha)^x(1+\beta)^y$$

for the function $\phi(x,y)$.

Now $\quad m_1\alpha + m_2\beta + \bar{m}\alpha\beta = (m_1-\bar{m})(\alpha+1) + (m_2-\bar{m})(\beta+1)$
$$+\bar{m}(\alpha+1)(\beta+1) + \bar{m} - \overline{m_1 + m_2},$$

so that picking out the coefficient of $(1+\alpha)^x(1+\beta)^y$ on the left hand side, we have as the sought frequency function

$$\phi(x,y) = e^{-(m_1+m_2-\bar{m})}\left\{\frac{(m_1-\bar{m})^x}{x!}\cdot\frac{(m_2-\bar{m})^y}{y!} + \frac{(m_1-\bar{m})^{x-1}}{(x-1)!}\cdot\frac{(m_2-\bar{m})^{y-1}}{(y-1)!}\cdot\frac{\bar{m}}{1!} + \cdots\right\}$$

$$= e^{-(m_1+m_2-\bar{m})}\sum_{v=0}^{\text{lesser } x,y} \frac{(m_1-\bar{m})^{x-v}}{(x-v)!}\cdot\frac{(m_2-\bar{m})^{y-v}}{(y-v)!}\cdot\frac{\bar{m}^v}{v!} \quad \ldots\ldots (16\cdot2).$$

§17. EXPANSION OF $\phi(x,y)$ IN A SERIES OF TYPE B. POLYNOMIALS.

The above form of the frequency function is not particularly convenient to work with, so we seek another, analogous to the second form of the normal correlation function.

If (16·1) is expanded as a power series in $\bar{m}$ then we have

$$\sum_{v=0}^{\infty} \alpha^v e^{m_1\alpha}\cdot\beta^v e^{m_2\beta}\,\frac{\bar{m}^v}{v!},$$

which by the result of §5 is equivalent to

$$\sum_{v=0}^{\infty}\left(\sum_x (-)^v K_v(x)\,\psi(x)\,(1+\alpha)^x\right)\left(\sum_y (-)^v K_v(y)\,\psi(y)(1+\beta)^y\right)\frac{\bar{m}^v}{v!},$$

where

$$\psi(x) = e^{-m_1}m_1^{\,x}/x!, \quad \psi(y) = e^{-m_2}m_2^{\,y}/y!,$$

and $K_v(x), K_v(y)$ are the Type B polynomials appropriate to these functions respectively. Thus finally, assuming that the order of the summations may be interchanged,

$$\phi(x,y) = \frac{e^{-m_1} m_1^x}{x!} \cdot \frac{e^{-m_2} m_2^y}{y!} \left(1 + K_1(x).K_1(y) \frac{\bar{m}}{1!} + K_2(x).K_2(y) \frac{\bar{m}^2}{2!} + \cdots\right) \cdots\cdots (17\cdot1).$$

This could also be written in the operational form

$$\phi(x,y) = P(\nabla_x, \nabla_y)\, \frac{e^{-m_1} m_1^x}{x!} \cdot \frac{e^{-m_2} m_2^y}{y!}$$

where $P(\alpha,\beta) = e^{\bar{m}\alpha\beta}$.

## 18. THE FACTORIAL MOMENTS OF THE DISTRIBUTION.

The factorial m.g.f. as found above is $e^{m_1\alpha + m_2\beta + \bar{m}\alpha\beta}$, so that if $m(r,s)$ is the double factorial moment of order $r$ in $x$, and $s$ in $y$, it is the coefficient of $\alpha^r\beta^s/r!\,s!$ in this, whence

$$m(r,s) = r!\,s! \sum_{v=0}^{\text{lesser of } r,s} \frac{m_1^{r-v}}{(r-v)!} \cdot \frac{m_2^{s-v}}{(s-v)!} \cdot \frac{\bar{m}^v}{v!} \cdots\cdots\cdots (18\cdot1)$$

In particular we have

$$m_{(1,0)} = \sum_x \sum_y x\,\phi(x,y) = m_1,$$

$$m_{(0,1)} = \sum_x \sum_y y\,\phi(x,y) = m_2$$

and

$$m_{(1,1)} = \sum_x \sum_y xy\,\phi(x,y) = m_1 m_2 + \bar{m}.$$

These three moments can be computed directly from any given data and so the three parameters in the frequency function are completely determined; also, further factorial moments of the distribution can be obtained by summing the necessary terms of (18·1).

The distribution of $x$ alone is obtained by summing (16·2) over all values of $y$ and is thus given by

$$e^{-m_1}\left(\frac{(m_1-\bar{m})^x}{x!} + \frac{(m_1-\bar{m})^{x-1}}{(x-1)!}\cdot\frac{\bar{m}}{1!} + \cdots + \frac{\bar{m}^x}{x!}\right)$$

$$= e^{-m_1} m_1^x / x!\,,$$

where, as found above, $m_1$ is the mean of the $x$'s. Similarly, the distribution of $y$ alone is given by $e^{-m_2} m_2^{\,y}/y!$ where $m_2$ is the mean of the $y$'s, so that the totals of the $x$- and $y$- arrays in the correlation table, are distributed according to the Poisson exponential law.

## §19. REGRESSION LINES.

The mean $x$ corresponding to a fixed $y = k$, which we shall dente by $\hat{x}_k$, is given by

$$\hat{x}_k = \sum_x x\phi(x,k) \Big/ \sum_x \phi(x,k).$$

$$\sum_x x\phi(x,k) = e^{-m_2}\left\{(m_1-\bar{m})\frac{(m_2-\bar{m})^k}{k!} + (m_1-\bar{m}+1)\frac{(m_2-\bar{m})^{k-1}}{(k-1)!}\cdot\frac{\bar{m}}{1!} + \dots\dots + (m_1-\bar{m}+k)\frac{\bar{m}^k}{k!}\right\}$$

$$= e^{-m_2}\left\{(m_1-\bar{m})\frac{m_2^{\,k}}{k!} + \bar{m}\cdot\frac{m_2^{\,k-1}}{(k-1)!}\right\}$$

and $\quad \sum_x \phi(x,k) = e^{-m_2} m_2^{\,k}/k!,$

so that $\quad \hat{x}_k = m_1 - \bar{m} + \frac{\bar{m}}{m_2}k.$

Hence the locus of the means of the $x$'s corresponding to any $y$ is

$$\hat{x} = m_1 - \bar{m} + \frac{\bar{m}}{m_2}y$$

or $\quad \hat{x} - m_1 = \frac{\bar{m}}{m_2}(y - m_2).$

Similarly the locus of the means of the $y$'s corresponding to any $x$ is given by

$$\hat{y} - m_2 = \frac{\bar{m}}{m_1}(x - m_1).$$

These lines may be regarded as the "regression lines" of the distribution although they will not, as in the case of normal correlation, given the most probable value of one variable corresponding to any value of the other.

We may further extend the analogy and define a coefficient of correlation, $r$ by

$$r = \sqrt{\frac{\bar{m}}{m_1} \cdot \frac{\bar{m}}{m_2}} = \frac{\bar{m}}{\sqrt{m_1 m_2}} = \frac{m_{(1,1)} - m_{(0,1)}\, m_{(1,0)}}{\sqrt{m_{(0,1)} \cdot m_{(1,0)}}}.$$

Since in the theoretical distribution being considered the totals of the $x$- and $y$-arrays form Poisson distributions, the variances of these are equal to the means of the distributions, so that the above result may be written,

$$r = \frac{\mu_{1,1}}{\sqrt{\mu_{0,2} \cdot \mu_{2,0}}}$$

where $\mu_{1,1}$ is the ordinary cross moment, $\mu_{2,0}$, $\mu_{0,2}$ the variances of the $x$- and $y$-arrays, all three being calculated from the means of the distribution. This is just the definition of the Pearsonian coefficient of correlation.

This definition can also be arrived at from another direction. In the case of Poisson correlation, the factor expressing the correlation involves the Type B. polynomials which contain the parameters $m_1$, and $m_2$. In the series, on the other hand, that we took as representing the correlation in the case of variables distributed normally, the variables are normalized so that the parameter $r$ is calculated in terms of absolute units. The true analogue in the normal case will therefore be the series expression containing the parameter $r$ and the Hermite polynomials, when these are calculated with respect to $e^{-\frac{1}{2}x^2/\sigma_x^2}$ and $e^{-\frac{1}{2}y^2/\sigma_y^2}$. Since

$$H_s(x, \sigma_x) = (-)^s \frac{d^s}{dx^s} e^{-\frac{1}{2}x^2/\sigma_x^2}$$

$$= \frac{1}{(\sigma_x)^s} H_s(x/\sigma_x),$$

the required form for normal correlation is

$$1 + r\sigma_x\sigma_y \cdot H_1(x,\sigma_x)\,H_1(y,\sigma_y) + \frac{r^2\sigma_x^2\sigma_y^2}{2!} H_2(x,\sigma_x)\,H_2(y,\sigma_y) + \cdots,$$

to which the form for Poisson correlation

$$1 + \bar{m}\,K_1(x,m_1).K_1(y,m_2) + \frac{\bar{m}^2}{2!} K_2(x,m_1)\,K_2(y,m_2) + \cdots$$

may be compared.

From the nature of the Poisson function and the Type B. polynomials we are compelled to leave the parameters $m_1$ and $m_2$ implicit, not being able to remove them by "normalizing" the variables, on account of the factorial expressions involved. When we considered the parameter $\bar{m}$ "normalized", so as to obtain its value in absolute units, we arrive at the definition for the correlation coefficient given above. It is known that the Pearsonian coefficient can be applied to distributions other than the normal, though, then, its distribution is not always known.

## §20. THE TYPE B. CORRELATION FUNCTION.

The distribution given by (16·2), can be derived under more general conditions. We shall now consider the samples of $N$ to be formed by taking one individual from each of $N$ different populations constituted similarly to the population taken in §16, that is, consisting of two correlated characters. If the parameters of the populations are $p_1, p_1', d_1;\ p_2, p_2', d_2; \cdots\cdots ; p_N, p_N', d_N$, then the factorial m.g.f. of the samples of $N$, since the selections are from independent populations, is

$$\prod_{r=1}^{N} (1 + p_r\alpha + p_r'\beta + d_r\alpha\beta)$$
$$= 1 + \Sigma p_r\alpha + \Sigma p_r'\beta + \Sigma d_r\alpha\beta$$
$$+ \Sigma p_r p_s \alpha^2 + \Sigma p_r' p_s' \beta^2 + \Sigma p_r p_s' \alpha\beta + \cdots,$$

where $\sum p_r p_s = \frac{1}{2}\{ (\sum p_r)^2 - \sum p_r^2 \}$

$$\sum p_r' p_s' = \frac{1}{2}\{ (\sum p_r')^2 - \sum p_r'^2 \}$$

$$\sum p_r p_s' = \sum p_r \sum p_r' - \sum p_r p_r'.$$

Let $\sum p_r = m_1$, $\sum p_r' = m_2$, and $\sum(d_r - p_r p_r') = \overline{m}$, which we assume all remain finite for large values of $N$. Thus the factorial m.g.f. may be written

$$e^{m_1 \alpha + m_2 \beta + \overline{m} \alpha\beta} \Big( 1 + B_{2,0}\alpha^2 + B_{0,2}\beta^2 + B_{3,0}\alpha^3 + B_{2,1}\alpha^2\beta + B_{1,2}\alpha\beta^2 + B_{0,3}\beta^3 + \cdots \Big); \quad \text{......} (20\cdot1).$$

where $B_{2,0}$ and $B_{0,2}$ are of order $1/N$, the other parameters being of successively higher order. In the limiting case we, therefore, have the same factorial m.g.f. as in §16, with a corresponding frequency function $\phi(x, y)$.

If we do not neglect the terms in $\alpha^2$ and $\beta^2$, and higher powers, then since $\alpha$ is equivalent to the operation $\Delta_x$ and $\beta$ to $\Delta_y$, we shall have, as the frequency function,

$$f(x,y) = \phi(x,y) + B_{2,0}\Delta_x^2 \phi(x,y) + B_{0,2}\Delta_y^2 \phi(x,y) + \cdots,$$

where $\phi(x,y)$ has the form given by (16·2). The function $f(x,y)$ may be called the correlation function of Type B. and compared with the Type A. correlation function (20),

$$\phi(x,y) + \sum_{k,r} A_{k,r} \frac{\partial^{k+r}}{\partial x^k \partial y^r} \phi(x,y),$$

where $k+r \nless 3$, and $\phi(x,y)$ is the normal correlation function in two variables.

## §21. THE PARAMETERS OF THE TYPE B. CORRELATION FUNCTION.

From (20·1), the factorial m.g.f. of the Type B. correlation function, we can obtain the necessary relations for the calculation of the parameters $B_{2,0}, B_{0,2}, \cdots$, the first of which will be given below.

$$m_1 = \text{coefficient of } \alpha = m_{(1,0)},$$
$$m_2 = \text{"} \quad \text{"} \ \beta = m_{(0,1)},$$
$$\overline{m}+m_1m_2 = \text{"} \quad \text{"} \ \alpha\beta = m_{(1,1)},$$
$$m_1^2+2B_{2,0} = \text{"} \quad \text{"} \ \alpha^2/2! = m_{(2,0)},$$
$$m_2^2+2B_{0,2} = \text{"} \quad \text{"} \ \beta^2/2! = m_{(0,2)},$$
$$m_1^3+6m_1B_{2,0}+6B_{3,0} = \text{"} \quad \text{"} \ \alpha^3/3! = m_{(3,0)}$$
$$m_2^3+6m_2B_{0,2}+6B_{0,3} = \text{"} \quad \text{"} \ \beta^3/3! = m_{(0,3)}$$
$$2m_1B_{02}+2B_{1,2} = \text{"} \quad \text{"} \ \alpha\beta^2/2! = m_{(1,2)}, \text{ etc.,}$$

where $m_{(r,s)}$ is the factorial moment of $f(x,y)$, of order $r$ in $x$ and $s$ in $y$.

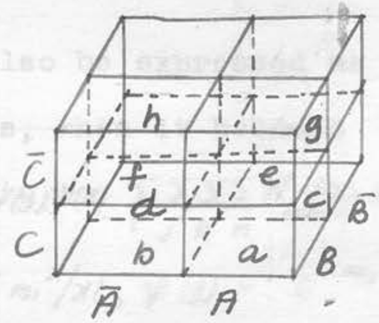
For $B_{s,0}$ and $B_{0,s}$ we shall have expressions in $m_{(r,0)}$ and $m_{(0,r)}$ $(r=1,2,\cdots,s_n)$ similar to (12·1). Thus

$$B_{2,0} = \tfrac{1}{2}\left(m_{(2,0)} - m_{(1,0)}^2\right),$$
$$B_{3,0} = \tfrac{1}{6}\left(m_{(3,0)} - 3m_{(2,0)}m_{(1,0)} + 2m_{(1,0)}^3\right),$$
$$B_{0,2} = \tfrac{1}{2}\left(m_{(0,2)} - m_{(0,1)}^2\right),$$
$$B_{0,3} = \tfrac{1}{6}\left(m_{(0,3)} - 3m_{(0,2)}m_{(0,1)} + 2m_{(0,1)}^3\right),$$

and
$$B_{1,2} = \tfrac{1}{2}\left(m_{(1,2)} - m_{(0,2)}m_{(1,0)} + m_{(0,1)}^2 m_{(1,0)}\right)$$
$$B_{2,1} = \tfrac{1}{2}\left(m_{(2,1)} - m_{(2,0)}m_{(0,1)} + m_{(1,0)}^2 m_{(0,1)}\right).$$

§22. <u>POISSON CORRELATION FUNCTION IN THREE VARIABLES.</u>

An extension of the result found earlier can be made to the case where the population consists of three correlated characters $A, B, C$, the distribution of which may be represented diagrammatically by the following eightfold table.




Let the relative frequency of $A$ in the population be $p_1$, that is $(a+c+e+g)$ (assuming total frequency is unity), and for the absence of $A$ be $\bar{p}_1$, and similarly $p_2, \bar{p}_2$ for $B$, and $p_3, \bar{p}_3$ for $C$.

The frequency generating function of the population is therefore

$$h + gA + fB + dC + eAB + bBC + cCA + aABC,$$

from which, precisely as before, we obtain the factorial m.g.f. of samples of $N$, as

$$\left(1 + p_1\alpha + p_2\beta + p_3\gamma + (a+e)\alpha\beta + (a+b)\beta\gamma + (a+c)\gamma\alpha + a\,\alpha\beta\gamma\right)^N.$$

Now let $Np_1 = m_1$, $Np_2 = m_2$, $Np_3 = m_3$,

$N(a+e-p_1p_2) = m_{1,2}$, $N(a+b-p_2p_3) = m_{2,3}$, $N(c+a-p_3p_1) = m_{3,1}$

and $N(a - p_1p_2p_3) = m_{1,2,3}$, concerning all of which we make the assumption that they remain finite as $N$ becomes larger. In the limit the factorial m.g.f. takes the form,

$$e^{m_1\alpha + m_2\beta + m_{1,2}\alpha\beta + m_{2,3}\beta\gamma + m_{3,1}\gamma\alpha + m_{1,2,3}\alpha\beta\gamma}.$$

The solution of this gives the frequency function $f(x,y,z)$ which can be represented in the operational form

$$f(x,y,z) = P(\nabla_x, \nabla_y, \nabla_z).e^{-(m_1+m_2+m_3)}\frac{m_1^x}{x!}.\frac{m_2^y}{y!}\frac{m_3^z}{z!},$$

where

$$P(\alpha,\beta,\gamma) = e^{m_{1,2}\alpha\beta + m_{2,3}\beta\gamma + m_{3,1}\gamma\alpha + m_{1,2,3}\alpha\beta\gamma}.$$

It may also be expressed as a series involving the Type B. polynomials, when it becomes

$$\phi(x,y,z) = \psi(x).\psi(y).\psi(z).\sum_i\sum_j\sum_k\sum_n K_{i+k+n}(x).K_{j+i+n}(y).K_{j+k+n}(z).\frac{m_{1,2}^i}{i!}\frac{m_{2,3}^j}{j!}\frac{m_{3,1}^k}{k!}\frac{m_{1,2,3}^n}{n!},$$

where $\psi(x) = e^{-m_1} m_1^x/x!$, $\psi(y) = e^{-m_2} m_2^y/y!$, $\psi(z) = e^{-m_3} m_3^z/z!$,

and $K_\nu(x)$ is formed with respect to $\psi(x)$, $K_\nu(y)$ with respect to $\psi(y)$ and $K_\nu(z)$ with respect to $\psi(z)$.

Finally from the factorial m.g.f. we obtain the relations connecting the parameters $m_1, m_2, \cdots, m_{1,2,3}$ with the factorial moments of the distribution.

$$m_1 = m_{(1,0,0)}, \quad m_2 = m_{(0,1,0)}, \quad m_3 = m_{(0,0,1)},$$

$$m_{1,2} = m_{(1,1,0)} - m_{(1,0,0)} m_{(0,1,0)},$$

$$m_{2,3} = m_{(0,1,1)} - m_{(0,1,0)} m_{(0,0,1)},$$

$$m_{3,1} = m_{(1,0,1)} - m_{(0,0,1)} m_{(1,0,0)},$$

$$m_{1,2,3} = m_{(1,1,1)} - m_{(1,0,0)} m_{(0,1,1)} - m_{(0,1,0)} m_{(1,0,1)} - m_{(0,0,1)} m_{(1,1,0)} + 2 m_{(1,0,0)} m_{(0,1,0)} m_{(0,0,1)},$$

where $m_{(r,s,t)}$ is the factorial moment of the frequency function $\phi(x,y,z)$ of order $r$ in $x$, $s$ in $y$, $t$ ~~and~~ in $z$.

§23. NUMERICAL EXAMPLE.

As it was stated in the introduction, no statistical data of two correlated variables, each following the Poisson law of distribution, could be found. The following experiment, suggested by those of A.D.Darbishire (5) for normal correlation, was devised.

The essential part of the experiment was the same as that by which the Poisson distribution in § 7 was obtained. It was repeated fifty times, thus giving in all 500 results. The distribution of these is given below.

| $x$ | $f(x)$ | | | | Calculated. |
|---|---|---|---|---|---|
| 0 | 296 | 500 | | | 303 |
| 1 | 165 | 204 | 250 | | 151 |
| 2 | 33 | 39 | 46 | 54 | 38 |
| 3 | 5 | 6 | 7 | 8 | 7 |
| 4 | 1 | 1 | 1 | 1 | 1 |

From this we have $m_{(1)} = \frac{250}{500} = \cdot 5$, $m_{(2)} = \frac{108}{500} = \cdot 22$, by which the values in the last column were obtained. For some reason, connected possibly both with the nature of the marbles used and with the method of performing the experiment, the distribution deviates from the expected Poisson distribution, particularly in the classes $x=0$, $x=1$, there being an excess in the class $x=1$.

These 500 results were then grouped in pairs, that is, if $u, v, w$ are three consecutive recordings of the original series, a new sequence was formed with terms like $u+v$, $v+w$, and so on. This was taken to be the $x$-series. The $y$-series was obtained by shifting the $x$-sequence one place to the left, which thus gave a set of 500 pairs of terms, the components of each pair being of the form $(u+v, v+w)$.

The $u, v, w$ are uncorrelated, and are subject, theoretically, to a Poisson Law of distribution with mean $\cdot 5$, and so with variance $\cdot 5$. Therefore we should expect to have

$$\sigma_x^2 = \sigma_u^2 + \sigma_v^2 = 1$$

$$\sigma_y^2 = \sigma_v^2 + \sigma_w^2 = 1,$$

$$r\sigma_x\sigma_y = \Sigma(u+v)(v+w)$$

$$= \sigma_v^2 \quad , \text{ since } u, v, w, \text{ are independent.}$$

Hence $r$ has the theoretical value $r = \cdot 5$ as might be expected.

From these pairs of terms the following correlation table was obtained.

| $x$ \ $y$ | 0 | 1 | 2 | 3 | 4 | 5 | $f_x$ | $xf_x$ | $x^2f_x$ | $yf_{x_y}$ | $\to xx$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 83 | 60 | 15 | 5 | | | 163 | | | 105 | |
| 1 | 64 | 103 | 37 | 8 | | 1 | 213 | 213 | 213 | 206 | 206 |
| 2 | 14 | 39 | 23 | 12 | 1 | | 89 | 178 | 356 | 125 | 250 |
| 3 | 1 | 7 | 14 | 4 | 3 | | 29 | 87 | 261 | 59 | 177 |
| 4 | 1 | 3 | 1 | | | | 5 | 20 | 80 | 5 | 20 |
| 5 | | | | | 1 | | 1 | 5 | 25 | 4 | 20 |
| $f_y$ | 163 | 212 | 90 | 29 | 5 | 1 | 500 | 503 | 935 | 504 | 673 |
| $yf_y$ | | 212 | 180 | 87 | 20 | 5 | 504 | | | | |
| $y^2f_y$ | | 212 | 360 | 261 | 80 | 25 | 938 | | | | |
| $xf_{y_x}$ | 99 | 214 | 129 | 44 | 16 | 1 | 503 | | | | |
| $yx\downarrow$ | | 214 | 258 | 132 | 64 | 5 | 673 | | | | |

$$m_{(1,0)} = \frac{503}{500} = 1\cdot 006. \qquad m_{(0,1)} = \frac{504}{500} = 1\cdot 008.$$

$$\sigma_x^2 = \frac{935}{500} - 1\cdot 01 = \cdot 86. \quad \sigma_y^2 = \frac{938}{500} - 1\cdot 01 = \cdot 86.$$

$$\overline{m} = \frac{675}{500} - 1\cdot 01 = \cdot 34.$$

We also have

$$m_{(2,0)} = \cdot 86. \qquad m_{(0,2)} = \cdot 87.$$

Finally

$$r = \frac{\overline{m}}{\sigma_x \sigma_y} = \cdot 40.$$

That the distribution deviates from the theoretical Poisson form can be seen from the table itself, and it is also brought out by the values of the variances and second factorial moments.

This deviation is no doubt largely due to the imperfectness of the original frequency distribution. It may also have another source. It is well known that the sum of two independent Poisson distributions is a Poisson - this can be seen at once, for if $e^{a_1\alpha}$ is the factorial m.g.f. of the first distribution, and $e^{a_2\alpha}$ of the second, the factorial m.g.f. of the sum, when the components are independent, is $e^{(a_1+a_2)\alpha}$, clearly a Poisson function. In our case, however, the $x$- and $y$- sequences are not really formed from independent distributions, as the consecutive terms in each sequence have a common element. It is not clear what will be the effect of this, but it may be that it introduces a further element tending to make the distributions more of the Type B. form.

Regarding the correlation table given above as a contingency table we now construct, using the end totals, a further table which gives the distribution of frequency that would be expected if the variables were not correlated.

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| 0 | 53 | 69 | 30 | 9 | 2 |
| 1 | 69 | 91 | 38 | 13 | 2 |
| 2 | 29 | 38 | 16 | 5 | 1 |
| 3 | 10 | 12 | 5 | 2 | |
| 4 | 2 | 2 | 1 | | |

It is formed in the usual way, the value in the class $(1,1)$, for example, is obtained from $\frac{212 \times 213}{500}$.

By the result (17.1) the effect of the correlation is expressed by the series

$$1 + K_1(x) K_1(y)\, \bar{m} + K_2(x) K_2(y) \frac{\bar{m}^2}{2!} + \cdots$$

in the frequency function. We shall now find the value of this for the various classes and with these, we shall be able to construct a frequency distribution from the above table, that may be compared with that in the original correlation table. For this purpose we make use of the tables of $K_r(x)$ to be found in the appendix. In this case, we shall require the table for $a=1$ as this is the value of the parameter in both sets of polynomials.

Since $m_{(1,0)} = m_{(0,1)} = 1$ we have here, theoretically $\bar{m} = r$, but as the data are not of the true Poisson form, these two values, when $r$ is evaluated by the second definition given in §19, do not coincide. From the method by which the data were obtained the values of $m_{(0,1)}$ and $m_{(1,0)}$ must take the theoretical value unity no matter what disturbing conditions may exist, so that the first definition of the correlation coefficient given in §19 fails to take account of any such disturbance as it contains only the first moments. Adopting, therefore, the second definition, the Pearsonian coefficient, which, as it contains the second moments, will take account of the 'skewness' of the data, we shall use the value ·4 for the parameter.

Thus for the class $(0,0)$ the factor is

$$1 + \cdot 4 + \frac{(\cdot 4)^2}{2!} + \frac{(\cdot 4)^3}{3!} + \cdots = e^{\cdot 4} = 1{\cdot}49.$$

Take now the class $(0,1)$, for which the factor will be the same as for $(1,0)$ from the symmetry; it is

$$1 - \frac{(\cdot 4)^2}{2!} - \frac{(\cdot 4)^3}{3!}2 - \frac{(\cdot 4)^4}{4!}3 - \cdots$$

$$= 1 - \cdot 08 - \cdot 021 - \cdot 003 \quad (\text{nearly}).$$

$$= \cdot 89.$$

For $(1,1)$ we have,

$$1+\frac{(\cdot 4)^2}{2!}+\frac{(\cdot 4)^3}{3!}\cdot 4+\frac{(\cdot 4)^4}{4!}\,9+\cdots$$
$$= 1+\cdot 08+\cdot 043+\cdot 010 \text{ (nearly)}.$$
$$= 1\cdot 13.$$

Similarly for $(0,2),(2,0)$ we find the value as $\cdot 54$ and for $(2,1),(1,2)$ as $1\cdot 04$.

When we apply these factors to the appropriate frequencies in the last table we obtain the following partial table which may be compared with the corresponding part of the correlation table.

| | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | (45) 49 | (64) 61 | (18) 16 | (3) 4 |
| 1 | (64) 61 | (99) 103 | (39) 39 | |
| 2 | (18) 16 | (39) 39 | | |
| 3 | (3) 4 | | | |

The figures in brackets are the frequencies corresponding to taking the value of the parameter in the correlation factor as $\cdot 34$.

The factor $1+K_1(x)K_1(y)\,\bar{m}+K_2(x)K_2(y)\frac{\bar{m}^2}{2!}+\cdots$ applied to the uncorrelated frequencies thus brings them into quite reasonable agreement with the observed correlated frequencies. The discrepancies in the classes $(0,0)$ and $(1,1)$ for the bracketed frequencies are undoubtedly a reflection of the discrepancies noted earlier, since the value $\cdot 34$ of the parameter does not take this deviation into account.

As far as we have been able to ascertain, the possibility of the existence of correlation between variables following a

Poisson Law of distribution has hitherto been denied. In a paper by Greiner (21. p.150) it is demonstrated that no correlation can exist between such variables. He obtains this result by considering the limit, under Poisson conditions, of

$$\frac{n!}{s_1!\,s_2!\,(n-s_1-s_2)!}\,p_1^{\,s_1}p_2^{\,s_2}(1-p_1-p_2)^{n-s_1-s_2},$$

which is

$$\frac{e^{-m_1}m_1^{\,s_1}}{s_1!}\cdot\frac{e^{-m_2}m_3^{\,s_2}}{s_2!}\,.$$

This is evidently sampling from the specialized population whose frequency generating function is

$$1-p_1-p_2+p_1t_1+p_2t_2\,,$$

or factorial m.g.f. $\quad 1+p_1\alpha+p_2\beta,$

which differs from the population we considered in the absence of the term $d\alpha\beta$ . While it is true that the corresponding treatment of the above population under normal conditions yields the normal correlation function, analogies between the normal and Poisson cannot be pressed and converse theorems do not always hold.

The data from the experiment discussed in this paragraph showed an effect due to correlation and the mathematical interpretation of it undoubtedly lay in a bracket factor involving the Type **B**. polynomials.

# - A P P E N D I X -

I. Tables of $K_\nu(x)$.

II. Bibliography.

## I. TABLES OF $K_r(x)$ .

These tables cover the range of values that are likely to be required. A single page is devoted to each value of the parameter $a$ and $a$ goes by steps of ·1 from ·5 to 1 and then by steps of ·5 to 5. Intermediate values of $a$ can be interpolated for. The polynomials are tabulated up to the sixth order and for $x$ ranging by units from $x=0$ to $x=10$ .

The values were calculated using the relation (13·4). Owing to limitations in the capacity of the machine used, the values could not all be obtained to the same number of decimal places. They are given correct to the last figure tabulated.

A partial check on the working was given from the relation $K_r(r+1) = -K_{r+1}(r)$ ; also occasional end-values were calculated independently.

$a = \cdot 5$



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | 1 | −3 | 5 | −7 | 9 | −11 |
| 2 | 1 | 3 | 1 | −13 | 33 | −61 | 97 |
| 3 | 1 | 5 | 13 | −7 | −71 | 269 | −635 |
| 4 | 1 | 7 | 33 | 71 | −127 | −441 | 2593 |
| 5 | 1 | 9 | 61 | 269 | 441 | −1711 | −2699 |
| 6 | 1 | 11 | 97 | 635 | 2593 | 2699 | −23231 |
| 7 | 1 | 13 | 141 | 1217 | 7673 | 28629 | 9157 |
| 8 | 1 | 15 | 193 | 2063 | 17409 | 105359 | 352705 |
| 9 | 1 | 17 | 253 | 3221 | 33913 | 279449 | 1617013 |
| 10 | 1 | 19 | 321 | 4739 | 59681 | 618579 | 4970401 |

$a = 0.6$.

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | ·6667 | -2·3333 | 4 | -5·6667 | 7·333 | -9 |
| 2 | 1 | 2·3333 | -·1111 | -7·6667 | 21 | -39·889 | 64·33 |
| 3 | 1 | 4 | 7·6667 | -8·2222 | 30·1111 | 135·11 | -334·56 |
| 4 | 1 | 5·6667 | 21 | 30·1111 | -84·93 | -115·81 | 1016·7 |
| 5 | 1 | 7·3333 | 39·8889 | 135·1111 | 115·81 | -823·53 | -141·6 |
| 6 | 1 | 9 | 64·3333 | 334·5556 | 1016·56 | 141·59 | -8376·9 |
| 7 | 1 | 10·6667 | 94·3333 | 656·2222 | 3246·93 | 8612·89 | -6961·0 |
| 8 | 1 | 12·3333 | 129·8889 | 1127·8889 | 7621·74 | 35670·60 | 79167·9 |
| 9 | 1 | 14 | 171 | 1777·3333 | 15141·00 | 99185·111 | 435843·9 |
| 10 | 1 | 15·6667 | 217·6667 | 2632·3333 | 26989·89 | 225360·11 | 1427425·1 |

$a = .7$



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | .4286 | -1.8571 | 3.2857 | -4.7143 | 6.143 | -7.57 |
| 2 | 1 | 1.8571 | -.6327 | -4.6735 | 14.061 | -27.53 | 45.1 |
| 3 | 1 | 3.2857 | 4.6735 | -7.385 | -12.64 | 72.9 | -191 |
| 4 | 1 | 4.7143 | 14.0612 | 12.644 | -54.84 | -17.4 | 434 |
| 5 | 1 | 6.1429 | 27.5306 | 72.907 | 17.41 | -409.1 | 285 |
| 6 | 1 | 7.5714 | 45.0816 | 190.895 | 434.02 | -284.8 | -3222 |
| 7 | 1 | 9 | 66.7143 | 384.102 | 1524.84 | 2815.3 | -5663 |
| 8 | 1 | 10.4286 | 92.4286 | 670.020 | 5719.71 | 18707.1 | 18468 |
| 9 | 1 | 11.8571 | 122.2245 | 1066.142 | 7548.40 | 40276.5 | 135957 |
| 10 | 1 | 13.2857 | 156.1020 | 1589.962 | 13640.64 | 94195.6 | 481184 |

$a = \cdot 8$.

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | ·25 | −1·5 | 2·75 | −4 | 5·25 | −6·5 |
| 2 | 1 | 1·5 | −·875 | −2·875 | 9·75 | −19·75 | 32·875 |
| 3 | 1 | 2·75 | 2·875 | −6·1563 | −4·625 | 41·188 | −115·25 |
| 4 | 1 | 4 | 9·75 | 4·625 | −35·4063 | 12·281 | 193·66 |
| 5 | 1 | 5·25 | 19·75 | 41·1875 | −12·2813 | −209·008 | 285·77 |
| 6 | 1 | 6·5 | 32·875 | 115·25 | 193·2813 | −285·466 | −1281·79 |
| 7 | 1 | 7·75 | 49·125 | 238·5313 | 769·9063 | 922·242 | −3425·03 |
| 8 | 1 | 9 | 68·5 | 422·75 | 1962·5625 | 5734·156 | 3491·78 |
| 9 | 1 | 10·25 | 91 | 679·625 | 4076·3125 | 18000·172 | 46497·95 |
| 10 | 1 | 11·5 | 116·625 | 1020·875 | 7474·4375 | 43477·125 | 181499·24 |

$a = \cdot 9$.

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | ·1111 | −1·2222 | 2·3333 | −3·4444 | 4·556 | −5·67 |
| 2 | 1 | 1·2222 | −·9753 | −1·7407 | 6·926 | −14·58 | 24·7 |
| 3 | 1 | 2·3333 | 1·7407 | −4·9918 | −·811 | 23·90 | −72·5 |
| 4 | 1 | 3·4444 | 6·9259 | ·8107 | −22·996 | 19·39 | 86·8 |
| 5 | 1 | 4·5556 | 14·5802 | 23·8971 | −19·393 | −108·36 | 216·1 |
| 6 | 1 | 5·6667 | 24·7037 | 72·4978 | 86·816 | −216·10 | −506·3 |
| 7 | 1 | 6·7778 | 37·2963 | 154·8433 | 409·028 | 266·20 | −1947·0 |
| 8 | 1 | 7·8889 | 52·3580 | 279·1641 | 1097·220 | 2538·56 | −172·3 |
| 9 | 1 | 9 | 69·8889 | 453·6904 | 2337·948 | 8634·16 | 16751·5 |
| 10 | 1 | 10·1111 | 89·8889 | 686·6532 | 4354·347 | 21622·63 | 74312·8 |

$a=1$

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | · | −1 | 2 | −3 | 4 | −5 |
| 2 | 1 | 1 | −1 | −1 | 5 | −11 | 19 |
| 3 | 1 | 2 | 1 | −4 | 1 | 14 | −47 |
| 4 | 1 | 3 | 5 | −1 | −15 | 19 | 37 |
| 5 | 1 | 4 | 11 | 14 | −19 | −56 | 151 |
| 6 | 1 | 5 | 19 | 47 | 37 | −151 | −185 |
| 7 | 1 | 6 | 29 | 104 | 225 | 34 | −1091 |
| 8 | 1 | 7 | 41 | 191 | 641 | 1159 | −887 |
| 9 | 1 | 8 | 55 | 314 | 1405 | 4364 | 6067 |
| 10 | 1 | 9 | 71 | 479 | 2661 | 11389 | 32251. |

$a = 1.5$



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | -·3333 | -·3333 | 1 | -1·6667 | 2·3333 | -3 |
| 2 | 1 | ·3333 | -·7778 | ·3333 | 1 | -3·2222 | 6·3333 |
| 3 | 1 | 1 | -·3333 | -1·2222 | 1·8889 | ·1111 | -6·5556 |
| 4 | 1 | 1·6667 | 1 | -1·8889 | -1·3704 | 6·4074 | -6·1111 |
| 5 | 1 | 2·3333 | 3·2222 | ·1111 | -6·4074 | 1·8408 | 19·5185 |
| 6 | 1 | 3 | 6·3333 | 6·5556 | -6·1111 | -19·5785 | 26·8817 |
| 7 | 1 | 3·6667 | 10·3333 | 19·2222 | 11·3704 | -39·8889 | -51·1921 |
| 8 | 1 | 4·3333 | 15·2222 | 39·8889 | 62·6296 | -1·9877 | -210·7477 |
| 9 | 1 | 5 | 21· | 70·3333 | 169 | 206·7774 | -219·6988 |
| 10 | 1 | 5·6667 | 27·6667 | 112·3333 | 356·5556 | 770·1111 | 608·4126 |

$\underline{a=2}$.



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | -.5 | . | .5 | -1 | 1.5 | -2 |
| 2 | 1 | . | -.5 | .5 | . | -1 | 2.5 |
| 3 | 1 | .5 | -.5 | -.25 | 1 | -1 | -.5 |
| 4 | 1 | 1. | . | -1. | .5 | 1.5 | -3.5 |
| 5 | 1 | 1.5 | 1 | -1 | -1.5 | 2.75 | 1 |
| 6 | 1 | 2 | 2.5 | .5 | -3.5 | -1 | 9.25 |
| 7 | 1 | 2.5 | 4.5 | 4.25 | -2.5 | -9.75 | 6.25 |
| 8 | 1 | 3 | 7 | 11 | 6 | -16 | -23 |
| 9 | 1 | 3.5 | 10 | 21.5 | 28 | -1 | -71 |
| 10 | 1 | 4 | 13.5 | 36.5 | 71 | 69 | -74 |

$a = 2.5$.



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | −·6 | ·2 | ·2 | −·6 | 1 | −1·4 |
| 2 | 1 | −·2 | −·28 | ·44 | −·28 | −·2 | 1 |
| 3 | 1 | ·2 | −·44 | ·104 | ·424 | −·76 | ·52 |
| 4 | 1 | ·6 | −·28 | −·424 | ·5904 | ·088 | −1·304 |
| 5 | 1 | 1 | ·2 | −·76 | −·088 | 1·269 | −1·093 |
| 6 | 1 | 1·4 | 1 | −·52 | −1·304 | 1·093 | 1·9526 |
| 7 | 1 | 1·8 | 2·12 | ·68 | −2·136 | −1·515 | 4·5758 |
| 8 | 1 | 2·2 | 3·56 | 3·224 | −1·048 | −5·787 | ·9398 |
| 9 | 1 | 2·6 | 5·32 | 7·496 | 4·1104 | −7·883 | −12·949 |
| 10 | 1 | 3 | 7·4 | 13·88 | 16·104 | ·3378 | −31·8682 |

$q=3$.

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | −·6667 | ·3333 | · | −·3333 | ·6667 | −1 |
| 2 | 1 | −·3333 | −·1111 | ·3333 | −·3333 | ·1111 | ·3333 |
| 3 | 1 | · | −·3333 | ·2222 | ·1111 | −·4444 | ·5556 |
| 4 | 1 | ·3333 | −·3333 | −·1111 | ·4074 | −·2592 | −·3333 |
| 5 | 1 | ·6667 | −·1111 | −·4444 | ·2592 | ·4197 | −·8518 |
| 6 | 1 | 1 | ·3333 | −·5556 | −·3333 | ·8518 | −·0124 |
| 7 | 1 | 1·3333 | 1 | −·2222 | −1·0741 | ·2963 | 1·6912 |
| 8 | 1 | 1·6667 | 1·8889 | ·7778 | −1·3704 | −1·4939 | 2·2838 |
| 9 | 1 | 2 | 3 | 1·6667 | −·3333 | −3·7778 | −·7040 |
| 10 | 1 | 2·3333 | 4·3333 | 4·6667 | 1·8889 | −4·3333 | −8·2597 |

$a = 3.5$

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | −·7143 | ·4286 | −·1429 | −·1429 | ·4286 | −·7143 |
| 2 | 1 | −·4286 | ·0204 | ·2245 | −·3061 | ·2245 | ·0204 |
| 3 | 1 | −·1429 | −·2245 | ·2430 | −·0496 | −·2128 | ·4053 |
| 4 | 1 | ·1429 | −·3061 | ·0496 | ·2270 | −·2836 | ·040 |
| 5 | 1 | ·4286 | −·2245 | −·2128 | ·2836 | ·0406 | −·446 |
| 6 | 1 | ·7143 | ·0204 | −·4053 | ·0404 | ·4458 | −·376 |
| 7 | 1 | 1 | ·4286 | −·3878 | −·4228 | ·3881 | ·388 |
| 8 | 1 | 1·2857 | 1 | −·0204 | −·8659 | −·2159 | 1·053 |
| 9 | 1 | 1·5714 | 1·7347 | ·8367 | −·8892 | −1·4529 | ·683 |
| 10 | 1 | 1·8571 | 2·6327 | 2·3236 | ·0670 | −2·7232 | −1·808 |

$a = 4$. 

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | -·75 | ·5 | -·25 | · | ·25 | -·5 |
| 2 | 1 | -·5 | ·125 | ·125 | -·25 | ·25 | -·125 |
| 3 | 1 | -·25 | -·125 | ·2188 | -·125 | -·0625 | ·25 |
| 4 | 1 | · | -·25 | ·125 | ·0938 | -·2188 | ·1563 |
| 5 | 1 | ·25 | -·25 | -·0625 | ·2188 | -·1016 | -·1719 |
| 6 | 1 | ·5 | -·125 | -·25 | ·1563 | ·1719 | -·3242 |
| 7 | 1 | ·75 | ·125 | -·3438 | -·0938 | ·3672 | ·0664 |
| 8 | 1 | 1 | ·5 | -·25 | -·4375 | ·25 | ·4844 |
| 9 | 1 | 1·25 | 1 | ·125 | -·6875 | -·2969 | ·8594 |
| 10 | 1 | 1·5 | 1·625 | ·875 | -·5625 | -·1563 | ·4141 |

$a=45$.

| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | -1 | 1 | -1 | 1 | -1 | 1 |
| 1 | 1 | -·7778 | ·5556 | -·3333 | ·1111 | ·1111 | -·3333 |
| 2 | 1 | -·5556 | ·2099 | ·0370 | -·1852 | ·2346 | -·1852 |
| 3 | 1 | -·3333 | -·0370 | ·1770 | -·1523 | ·0288 | ·1276 |
| 4 | 1 | -·1111 | -·1852 | ·1523 | ·0050 | -·1404 | ·1660 |
| 5 | 1 | ·1111 | -·2346 | ·0288 | ·1404 | -·1348 | -·0212 |
| 6 | 1 | ·3333 | -·1852 | -·1276 | ·1660 | ·0212 | -·2009 |
| 7 | 1 | ·5556 | -·0370 | -·2510 | ·0526 | ·2056 | -·1726 |
| 8 | 1 | ·7778 | ·2099 | -·2757 | -·1705 | ·2640 | ·1015 |
| 9 | 1 | 1 | ·5556 | -·1358 | -·4156 | ·0746 | ·4536 |
| 10 | 1 | 1·2222 | 1 | ·2346 | -·5363 | -·3873 | ·5530 |

$q=5$



| $x$ | $K_0(x)$ | $K_1(x)$ | $K_2(x)$ | $K_3(x)$ | $K_4(x)$ | $K_5(x)$ | $K_6(x)$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | −1 | 1 | −1 | 1 | −1 | 1 |
| 1 | 1 | −·8 | ·6 | −·4 | ·2 | · | −·2 |
| 2 | 1 | −·6 | ·28 | −·04 | −·12 | ·2 | −·2 |
| 3 | 1 | −·4 | ·04 | ·128 | −·152 | ·08 | ·04 |
| 4 | 1 | −·2 | −·12 | ·152 | −·0496 | −·072 | ·136 |
| 5 | 1 | · | −·2 | ·08 | ·072 | −·1216 | ·0496 |
| 6 | 1 | ·2 | −·2 | −·04 | ·136 | −·0496 | −·0963 |
| 7 | 1 | ·4 | −·12 | −·16 | ·104 | ·0864 | −·1558 |
| 8 | 1 | ·6 | ·04 | −·232 | −·024 | ·1904 | −·0522 |
| 9 | 1 | ·8 | ·28 | −·208 | −·2096 | ·1664 | ·1763 |
| 10 | 1 | 1 | ·6 | −·04 | −·376 | −·0432 | ·376 |

## II. BIBLIOGRAPHY.


1. T.Kameda: Journ.Fac.Science Imp.University, Tokyo. Section I, Vol.I.(1925)(p.1-62.)

2. J.V.Uspensky: Annals of Math.Sec.Series 32.2 (1931) (p.306- )

3. J.V.Uspensky: Annals of Math. 28 (1927) (p.593 - )

4. H.Pollazcek-Geiringer: Skand. Aktuarietidskrift Haft, (1-2) (1928). p.(98-111).

5. A.D.Darbishire: "Calculus of Observations" (Whittaker and Robinson) p.330.

6. H.E.Soper: "Frequency Arrays". Cambridge Press (1922).

7. A.C.Aitken: Quarterly Journal of Mathematics. Vol.2.6 (1931). p. (130-135).

8. C.V.L.Charlier: Arkiv.för Math., Astron., Och Fysik. (1905-06) Bd.2.8.

9. H.Cramér: Skand. Aktuarietidskrift Haft (1-2) p.(13-97) and 3 (141-180).

10. C.V.L.Charlier; Medd. från Lunds Astron. Observ.II. 51. p.3.

11. W.Palin Elderton: "Frequency Curves and Correlation" 2nd Ed. London (1927).

12. W.F.Sheppard: Pro.Lond.Math.Soc. Series II. 13 (1913) p.81.

13. J.F.Steffensen: "Interpolation". Baltimore (1927).

14. T.N.Thiele: "Theory of Observations". London (1903).

15. Ramanujan: Proc.Lond.Math.Soc.Series II.19 (1921) (Hardy's Obituary) xlvii.

16. C.V.L.Charlier: Arkiv för Math.Astron.och Fysik Bd.2.15.

17. H.Pollazceck-Geiringer: Zeitschrift für Math.und Meck. 8.(1928) p.(292-309).

18. Ch.Jordan: Bulletin de la Soc.Math.de France. T.LIV-LV (1926-27) p.(101-37).

19. A.C.Aitken: Proc.Royal Soc.Edinburgh. LII.2. p.(174-182).

20. C.V.L.Charlier: Arkiv. för Math.Astron.och Fysik.
Bd.9 (1914). 26.

21. R.Greiner: Zeitschrift für Math. und Physik. Bd.57.
(1909). (p.121-158).